ESTIMATION METHODS FOR REGRESSION MODELS WITH UNEQUAL ERROR VARIANCES

GHULAM RASUL PASHA

THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF WARWICK
DEPARTMENT OF STATISTICS

JULY 1982

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

As is true of any such intellectual effort, the fundamental prerequisite for success is a mentor who embellishes the intellectual strengths and moral fibre of a student. Professor Keith Ord is a prototype of an ideal mentor. His brilliance is only matched by his tolerance for my shortcomings. I would like to thank him especially.

Thanks are also due to Professor P. J. Harrison, Chairman, Department of Statistics, Warwick University, and Professor George Heitman, Chairman, Division of Management Science, The Pennsylvania State University, USA, for their support and encouragement. I would also like to thank the Centre for Research, College of Business Administration, The Pennsylvania State University, for providing the computer time and Mrs. Bonnie K. Henninger for typing.

I would like to thank my wife, who has endured my whimsical moods and prolonged separation without the slightest hint of a complaint.

Finally, I would like to humbly dedicate this dissertation to my supervisor, Professor Keith Ord, a helpful and patient mentor and to my mother whose love and affection gave me the emotional strength to complete this work.

ABSTRACT

In this dissertation, we consider estimation procedures for the linear model when the observations are replicated but the error variances are unequal.

The aim of this research was to consider several standard methods for parameter estimation, namely Ordinary Least Squares, Weighted Least Squares, and Maximum Likelihood and to compare these with newly developed procedures. These new techniques included the use of a prior likelihood function to induce "shrinkage" towards a common value among the estimators for the error variances and procedures based on preliminary tests of the hypothesis of variance equality. Both an overall test of equality and a multiple comparison method were considered. In addition, variance estimates based on MINQUE (Minimum Norm Quadratic Unbiased Estimator) were investigated. The MINQU estimators tend to "stretch out" the variances and were found to be unsatisfactory.

The performance of the above-mentioned approaches was investigated both through asymptotic theoretical results and small samples simulation studies. The results from these two approaches were found to be in broad agreement.

Overall, the multiple comparison and prior likelihood procedures appear to perform best, but the prior likelihood depends upon the availability of satisfactory prior information. So the multiple comparison procedure appears to be the most effective technique in general.

A further study was also conducted to examine the effects of "errors-in-variables." In general it is found that maximum likelihood is superior to ordinary least squares and weighted least squares only if a large number of replicates is used.

CHAPTER 1

INTRODUCTION AND REVIEW OF EXISTING TECHNIQUES

1.1 Introduction

In the usual analysis of variance model, a typical observation is considered to be the result of some fixed effects and an error term combined in a linear fashion. In such models, interest mainly lies in estimating linear functions of the effects. There are, however, situations where in addition to estimating the linear functions of effects, the main interest is also concerned with the estimation of the variances of the effects. Effects of this nature are called random effects and variances associated with them are called variance components. A linear model in its generality may, of course, include some fixed and some random effects. Such a model is referred to as a variance components model. A linear regression model with a diagonal covariance matrix is a special case of the variance components model. In such a model, estimation of the unequal variances is our main goal as a way to generate improved estimators for the parameters describing the effects.

There is a vast literature on the topic of estimation of variance components. The usual mixed linear model discussed in the literature on variance components is

$$\underline{Y} = X\underline{\beta} + u_1\underline{\xi}_1+\cdots+u_k\underline{\xi}_k, \tag{1.1.1}$$

where Y is an n-vector of observations, X is an $(n\times p)$ known matrix, $\underline{\beta}$

is a fixed unknown p-vector of parameters, u_i is a given $(n \times n_i)$ matrix and ξ_i is an n_i-vector such that

$$E(\underline{\xi}_i) = \underline{0}, \ D(\underline{\xi}_i) = \sigma_i^2 I_{n_i} \quad \text{and} \quad Cov(\underline{\xi}_i, \underline{\xi}_j) = 0, \ i \neq j. \qquad (1.1.2)$$

The unknown parameters σ_i^2, $i=1,2,\ldots,k$ are called variance components.

A systematic study of the estimation of variance components was undertaken by Henderson (1953) who proposed three methods of estimation. But some of the early users of such models are due to Cochran (1939), Yates and Zacopancy (1935), Fairfield Smith (1936), Yates (1940), Panse (1946), Rao (1947, 1953, 1956), Henderson (1950) and Brownlee (1953) in different fields of applications.

The general approach in all these papers was to obtain k quadratic functions of $\underline{Y}$, say $\underline{Y}'A_i\underline{Y}$, $i=1,2,\ldots,k$ which are invariant for translation of Y by $X\alpha$ where α is arbitrary, and solve the equations.

$$\underline{Y}'A_i\underline{Y} = E(\underline{Y}'A_i\underline{Y}) = \sum_{i=1}^{k} \ell_i \hat{\sigma}_i^2 \qquad (1.1.3)$$

The method of choosing the quadratic forms was intuitive in nature (see Henderson, 1953) and did not depend on any stated criteria of estimation. The entries in the ANOVA table giving the sums of squares due to different effects were considered as good choices of the quadratic forms in general. The ANOVA technique provides good estimators in what are called balanced designs (see Anderson, 1975; Anderson and Crump, 1967) but, as shown by Seely (1975), such estimators may be inefficient in more general linear models. For a general discussion of Henderson's methods and their advantages (computational simplicity) and limitations (lack of uniqueness, inapplicability and inefficiency in special cases), see the papers by Searle (1968, 1971), Seely (1975), Olsen et al. (1976), and Harville (1977).

Hartley and Rao (1967) initiated a different approach in the maximum likelihood (ML) method. They considered the likelihood of the unknown parameters $\underline{\beta}$, σ_i^2, $i=1,2,\ldots,k$, based on observed Y and obtained the likelihood equations by computing the derivatives of the likelihood w.r. to the parameters. Patterson and Thompson (1975) considered another approach the marginal likelihood or the maximal invariant of $\underline{Y}$, i.e., only on C'Y where $C = X^{\perp}$ (matrix orthogonal to X) and obtained what are called marginal maximum likelihood (MML) equations. Harville (1977) has reviewed the ML and MML methods and the computational algorithms associated with them.

Maximum likelihood estimators, though consistent, may be heavily biased in small samples so that some caution is needed when they are used as estimates of individual parameters. The problem is not acute if the exact distribution of the ML estimator is known, since in that case, appropriate bias adjustments can be made in the individual estimators before using them. The general large sample properties associated with ML estimators are misleading in the absence of studies on the orders of sample sizes for which these properties hold in particular cases. The bias in the MML estimators may be slight even in small samples. As observed earlier, the MML estimator is, by construction, a function of C'Y the maximal invariant of Y. It turns out that even the full ML estimator is a function of C'Y although the likelihood is based on Y. There are important practical cases where reduction of Y to C'Y result in the non-identifiability of individual parameters, in which case, neither the ML nor the MML is applicable.

C. R. Rao (1970, 1971a, 1971b, 1972, 1973) proposed a new principle of estimating heteroscedastic variances and covariances components called MINQUE (minimum norm quadratic unbiased estimation, estimator or estimate, depending on context) the scope of which has been extended to cover a variety of situations by Focke and Dewess (1972), Kleffe (1975, 1976, 1977a, 1977b, 1978, 1979), J. N. K. Rao (1973), Fuller and J. N. K. Rao (1978), P. S. R. S. Rao and Chaubey (1978), P. S. R. S. Rao (1977), Pukelsheim (1977, 1978a), Sinha and Wieand (1977), and Rao (1979).

1.2 The MINQUE Principle

Consider the variance component model

$$\underline{Y} = X\underline{\beta} + u_1\underline{\xi}_1 + \cdots + u_k\underline{\xi}_k \tag{1.2.1}$$

where Y is an n-vector of observations, X is an $(n\times p)$ known matrix, β is a p-vector of parameters, u_i is a given $(n\times n_i)$ matrix and ξ_i is an n_i-vector such that

$$E(\xi_i) = 0,\ D(\xi_i) = \sigma_i^2 I_{n_i} \quad \text{and} \quad Cov(\underline{\xi}_i, \underline{\xi}_j) = 0,\ i\neq j. \tag{1.2.2}$$

We can express the above model in a compact form

$$\underline{Y} = X\underline{\beta} + U\underline{\xi} \tag{1.2.3}$$

where $U = (u_1|\ldots|u_k)$ and $\underline{\xi}' = (\underline{\xi}_1'|\ldots|\underline{\xi}_k')$.
From (1.2.2) we have

$$E(\underline{Y}) = X\underline{\beta}$$

$$\text{and } D(Y) = \sigma_1^2 v_1 + \sigma_2^2 v_2 + \ldots + \sigma_k^2 v_k \tag{1.2.4}$$

where $v_i = u_i u_i'$.

C. R. Rao (1971a) proposed to estimate a linear function $p_1\sigma_1^2+\ldots+p_k\sigma_k^2$ of the variance components $\sigma_1^2,\ldots,\sigma_k^2$ by a quadratic function Y'AY of random variables Y. He developed the following criteria for determining the matrix A.

1.2.1 Invariance Under Translation of the β Parameters. Instead of β, consider $\gamma = \beta-\beta_0$ as the unknown parameter, where β_0 is fixed. Model (1.2.3) becomes

$$\underline{Y} - X\underline{\beta}_0 = X\gamma+U\underline{\xi} \tag{1.2.5}$$

in which case the estimator of $\sum_i p_i\sigma_i^2$ is

$$(\underline{Y}-X\underline{\beta}_0)'A(\underline{Y}-X\underline{\beta}_0). \tag{1.2.6}$$

But (1.2.6) should have the same numerical value as Y'AY whatever β_0 may be. Thus, we require that

$$\begin{aligned}\underline{Y}'A\underline{Y} &= (\underline{Y}-X\underline{\beta}_0)'A(\underline{Y}-X\underline{\beta}_0)\\ &= \underline{Y}'A\underline{Y} + 2\underline{\beta}_0'X'A\underline{Y} + \underline{\beta}_0'X'AX\underline{\beta}_0 .\end{aligned} \tag{1.2.7}$$

The above equation is satisfied if

$$AX = 0. \tag{1.2.8}$$

1.2.2 Unbiasedness. With restriction (1.2.8)

$$\underline{Y}'A\underline{Y} = \underline{\beta}'X'AX\underline{\beta} + 2\underline{\beta}'X'AU\underline{\xi} + \underline{\xi}'U'AU\underline{\xi}$$

which reduces to

$$\underline{Y}'A\underline{Y} = \xi'U'AU\xi . \tag{1.2.9}$$

For $\underline{Y}'A\underline{Y}$ to be unbiased for $\sum_i p_i\sigma_i^2$ for all σ_i^2, we require

$$E(\underline{Y}'A\underline{Y}) = E(\xi'U'AU\xi) = \sum_i p_i\sigma_i^2 \tag{1.2.10}$$

since

$$E(\xi'U'AU\,\xi) = \sum_{i=1}^{k} E(\xi_i'u_i'Au_i\xi_i)$$

$$= \sum_{i=1}^{k} tr(u_i'Au_iE[\xi_i\xi_i'])$$

$$= \sum_{i=1}^{k} \sigma_i^2\ tr(u_i'Au_i)$$

$$= \sum_{i=1}^{k} \sigma_i^2\ tr\ Av_i \tag{1.2.11}$$

Here $\underline{tr}$ stands for trace and $v_i = u_iu_i'$. Equation (1.2.10) becomes

$$\sum_{i=1}^{k} \sigma_i^2\ tr\ Av_i = \sum_{i=1}^{k} p_i\sigma_i^2 \tag{1.2.12}$$

which implies that

$$tr\ Av_i = p_i \qquad i=1,2,\ldots,k\ . \tag{1.2.13}$$

1.2.3 Minimum Norm. If the hypothetical variables ξ were known, a natural unbiased estimator of $\sum_i p_i\sigma_i^2$ would be

$$\frac{p_1}{n_1}\,\xi_1'\xi_1 + \ldots + \frac{p_k}{n_k}\,\xi_k'\xi_k = \xi'\Delta\xi \tag{1.2.14}$$

where $\Delta = \text{diag}\left(\frac{p_1}{n_1} I_{n_1}, \ldots, \frac{p_k}{n_k} I_{n_k}\right)$.

But the proposed estimator is $\xi' U' A U \xi$. Hence we would like to choose A such that the difference $\xi'(U'AU-\Delta)\xi$ is minimized. Since ξ is unknown, Rao suggests that A is chosen to minimize

$$||U'AU-\Delta|| ,$$

where $||\cdot||$ denotes the norm of a matrix. In particular, with the Euclidean norm, the matrix A of the quadratic form Y'AY should be determined by minimizing

$$||U'AU-\Delta||^2 = \text{tr}(U'AU-\Delta)^2 \tag{1.2.15}$$

subject to (1.2.8) and (1.2.13). Under the condition in (1.2.13)

$$\begin{aligned}
\text{tr } U'AU\Delta = \text{tr } AU\Delta U' &= \text{tr } A\sum_i \frac{p_i}{n_i} u_i u_i' \\
&= \sum_i \frac{p_i}{n_i} \text{tr } Av_i \\
&= \sum_i \frac{p_i^2}{n_i} = \text{tr } \Delta^2
\end{aligned} \tag{1.2.16}$$

and hence

$$\begin{aligned}
\text{tr}(U'AU-\Delta)^2 &= \text{tr } U'AUU'AU - 2 \text{ tr } U'AU\Delta + \text{tr } \Delta^2 \\
&= \text{tr } U'AUU'AU - \text{tr } \Delta^2 \\
&= \text{tr } AVAV - \text{tr } \Delta^2
\end{aligned} \tag{1.2.17}$$

where $V = UU' = \sum_i u_i u_i' = \sum_i v_i$. Thus minimizing $\mathrm{tr}(U'AU-\Delta)^2$, Subject to (1.2.8) and (1.2.13) is equivalent to minimizing tr AVAV − tr Δ^2 or equivalently minimizing tr AVAV, since Δ is given, subject to (1.2.8) and (1.2.13). Thus, the principle of estimation may be described as follows.

The quadratic form $\underline{Y}'A\underline{Y}$ is said to be the MINQUE (Minimum Norm Quadratic Unbiased Estimator) of $\sum_i p_i\sigma_i^2$ where the matrix A is determined such that tr AVAV is minimized subject to the conditions (1.2.8) and (1.2.13).

C. R. Rao (1971a, 1972) also proposed the MINQUE principle without invariance which may be stated as follows. The quadratic form $\underline{Y}'A\underline{Y}$ is said to be the MINQUE without invariance of $\sum_i p_i\sigma_i^2$ if the matrix A is determined such that tr $A(V + \frac{1}{2}XX')A(V + \frac{1}{2}XX')$ is minimized subject to the conditions

$$X'AX = 0 \quad \text{and} \quad \mathrm{tr}\, Av_i = p_i, \quad i=1,2,\ldots,k. \tag{1.2.18}$$

<u>1.2.4 Examples of MINQUE</u>. C. R. Rao (1970) derives the MINQUE explicitly for some cases.

(i) When $E(\underline{Y}) = X\underline{\beta}$ and $D(\underline{Y}) = \sigma^2 I$, the MINQUE is the same as the usual Gauss-Markov estimator.

(ii) If $E(y_{ij}) = \alpha_i + \beta_j$ and $V(\epsilon_{ij}) = \sigma_i^2$, he shows that the MINQUE coincides with the estimator of Anscombe and Tukey (1963).

(iii) When $E(y_i) = \mu$ and $V(\epsilon_i) = \sigma_i^2$, $(i=1,\ldots,n)$, he obtains the MINQUE as

$$\hat{\sigma}_i^2 = \frac{n}{n-2}(y_i-\bar{y})^2 - \frac{s^2}{n-2} \tag{1.2.19}$$

where $\bar{y} = (\Sigma y_i/n)$ and $(n-1)s^2 = \Sigma(y_i-\bar{y})^2$.

Note: An intuitively appealing but biased set of estimators for this problem would be $\tilde{\sigma}_i^2 = \frac{n}{(n-1)}(y_i-\bar{y})^2$; (1.2.20)

but $E(\tilde{\sigma}_i^2) = \left(\frac{n-2}{n-1}\right)\sigma_i^2 + \frac{1}{n-1}\ \bar{\sigma}^2$.

By contrast with MINQUE which "stretches out" the estimators, (1.2.20) provides an element of "shrinkage" towards the overall mean $\bar{\sigma}^2 = \frac{1}{n}\Sigma\sigma_i^2$.

(iv) For the one-way random effects model with $E(y_{ij}) = \mu$ and $V(y_{ij}) = \sigma_\alpha^2 + \sigma^2$, $(i=1,2,\ldots,k;j=1,2,\ldots,m)$, we have verified that the MINQUE for σ^2 and σ_α^2 are the same as the usual ANOVA estimators.

(v) Now consider the multivariate model

$$Y_i = X_i\beta_i + \epsilon_i \qquad i=1,2,\ldots,p \tag{1.2.21}$$

where Y_i is an n vector, X_i is a known $(n\times m_i)$ matrix, β_i is an m_i vector and ϵ_i is an n-vector with mean zero and $E(\epsilon_i\epsilon_j') = \sigma_{ij}I$. For the restricted form the model is written as

$$\underline{Y} = X\beta + \underline{\epsilon} \tag{1.2.22}$$

where $Y' = (Y_1'|Y_2'|\ldots|Y_p')$, $\epsilon' = (\epsilon_1'|\epsilon_2'|\ldots|\epsilon_p')$, X and β are diagonal matrices with elements X_i and β_i respectively. Writing $D(\underline{\epsilon})$ in the form of (1.2.4), we obtain the MINQUE of σ_{ij} as $\hat{\sigma}_{ij} = e_ie_j/\text{tr}\ Q_iQ_j$ where $Q_i = I - X_i(X_i'X_i)^- X_i'$ and $e_i = Q_iy_i$. Zellner and Huang (1962) obtained this estimator for p=2 through a different approach. For the "unrestricted form," the model is written as

$$Y = (I \otimes X)\beta + \epsilon \tag{1.2.23}$$

where $\otimes$ denotes the Kronecker product of two matrices; and Y and ϵ are as above, but $X = (X_1|X_2|\ldots|X_p)$ and $\beta' = (\beta_1'|\beta_i'|\ldots|\beta_p')$. The MINQUE can now be shown to be $\hat{\sigma}_{ij} = e_i'e_j/\text{tr } Q$ where $Q = I - X(X'X)^{-}X'$ and $e_i = QY$. This estimator is the same as the one obtained by Zellner (1962).

<u>1.2.5 Properties of MINQUE</u>. (i) Additivity: If S_1 and S_2 are the MINQUE of $p_1'\sigma$ and $p_2'\sigma$ respectively, then (S_1+S_2) is the MINQUE of $(p_1+p_2)'\sigma$.

(ii) Invariance: Consider a non-singular transformation $Z = BY$. Clearly, $E(Z) = B\otimes\beta = X_*\beta$ and

$$D(Z) = \sum_{i=1}^{k} \sigma_i^2 B'v_i B = \sum_{i=1}^{k} \sigma_i^2 L_i \tag{1.2.24}$$

Now, the estimator of $p'\sigma$ is $Z'CZ$, where $A = B'CB$. C is obtained by minimizing tr CLCL subject to the conditions $CX_* = 0$ and tr $CL_i = p_i$, where $L = \sum_{i=1}^{k} L_i$. Here B is an orthogonal matrix. Rao (1971a) further suggested that by transforming to $Z = G'Y$, where G is an $n\times(n-r)$ such that $G'X = 0$, the computation can be further simplified.

(iii) Minimum Variance: When the elements of ξ_i have a common variance σ_i^2 and a common fourth moment μ_{4i}, the variance can be written as

$$V(\underline{Y}'A\underline{Y}) = \sum_{i=1}^{k} \sigma_i^4\gamma_i \text{ tr } Av_iAv_i + 2 \text{ tr } AD(Y)AD(Y) \tag{1.2.25}$$

where $\gamma_i = (\mu_{4i}/\sigma_i^4) - 3$.

If ξ_i are normally distributed, the first term on the right-hand side of (1.2.25) vanishes and, as Rao (1971b) points out, MINQUE coincides with the MIVQUE (Minimum Variance Quadratic Unbiased Estimator).

1.2.6 Modifications and Extensions. C. R. Rao (1970) first proposed the MINQUE principle for estimating heteroscedastic variances in a linear model. He also proposed using the estimated variances obtained through MINQUE to carry out the Weighted Least Squares (WLS) procedure.

In a Monte Carlo study, J. N. K. Rao and Subrahmaniam (1971) used MINQUE for two regression models and compares the WLS estimators of regression parameters using these MINQUE's with some other estimators of unequal variances. However, they ignored those samples which produced at least one negative estimate of variance in their Monte Carlo experiment for one of the models. On the same lines as Rao and Subrahmaniam (1971), Chaubey and Rao (1976) considered the models

$$Y_{ij} = \mu + \epsilon_{ij} \qquad (1.2.26)$$

$$Y_{ij} = \alpha + \beta x_i + \epsilon_{ij} \quad (i=1,\dots,k;j=1,2,\dots,n_i) \qquad (1.2.27)$$

where $\epsilon_{ij} \sim N(0,\sigma_i^2)$. They choose the patterns for σ_i^2 from Cochran and Carrol (1953). Efficiencies of the MINQUE, Average of the Squared Residuals (ASR) and the sample variance S_i^2 were compared. When the MINQUE took the negative values, it was replaced by S_i^2 or by a small positive quantity. They examined the effects of the above estimators on the WLS of α and β. J. N. K. Rao (1973) also evaluated the efficiencies of the above approaches for the variance estimators. Comparison between AUE (Almost-Unbiased Estimators) and the above estimators are made by Horn and Horn (1975) and Rao and Chaubey (1978).

For the model in (1.2.3)-(1.2.4), when the conditions $AX = 0$ and $\mathrm{tr}\, Av_i = p_i$ are not consistent, Focke and Dewess (1972) replaced the first condition with $X'AX = 0$. The resulting principle for estimating $p'\sigma$ given by Rao (1971a) as minimizing $\mathrm{tr}(V+XX')AV$ subject to $X'AX = 0$ and $\mathrm{tr}\, Av_i = p_i$. He derives the solution for A and an alternative form is given by Pringle (1974) as $A = \Sigma\lambda_i R_i$ and

$$\Sigma\lambda_i \;\mathrm{tr}\; R_i v_j = p_j, \tag{1.2.28}$$

where $R_i = (V + \frac{1}{2}XX')^{-1}(v_i - Pv_i P')(V + \frac{1}{2}XX')^{-1}$ and P is the projection operator $X(X'V^{-1}X)^{-}X'V^{-1}$. With A in (1.2.28), $Y'AY$ is unbiased for $P'\sigma$ but does not have the translation invariance unless $X'AX = 0$ implies $AX = 0$.

1.2.7 Some Merits and Drawbacks of MINQUE. The principle of MINQUE seems to be as fundamental in nature as the LS and ML methods of estimation.

The MINQUE is based on the LS residuals and possesses several appealing optimum properties.

(a) One or more restrictions such as invariance, unbiasedness and nonnegative definiteness can be placed on $Y'AY$ depending on the desired properties of the estimators.

(b) For a suitable choice of the norm, MINQU estimators provide minimum variance estimators when Y is normally distributed.

Horn, Horn, and Duncan (1975) mentioned the following deficiencies of the MINQUE.

(a) The MINQUE estimates σ_i^2 although unbiased may be negative.

(b) The MINQUE estimators require the inversion of an n×n matrix, i.e., the computations needed for obtaining MINQUE are somewhat difficult.

(c) The MINQU estimators do not exist for some models of interest.

Recently, some numerical techniques for computing MINQUE's have been tried by Ahrens (1978), Swallow and Searl (1978), Ahrens, et al. (1979), Infante (1978), and Kleffe (1980).

1.3 Maximum Likelihood Estimation (MLE)

1.3.1 The General Model. Consider the general model

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon},$$
$$E(\underline{\epsilon\epsilon'}) = \sum_{i=1}^{k} \theta_i v_i = V \tag{1.3.1}$$

and the ML estimation of θ under the assumption

$$\underline{Y} \sim N(X\underline{\beta},V),\ \underline{\beta}\epsilon R^m,\ \theta\epsilon F(\text{open set}) \tag{1.3.2}$$

we assume that V is p.d. for $\forall\theta\epsilon F$.

Harville (1977) has presented a review of the ML estimation of θ describing the contributions made by Hartley and Rao (1967), Anderson (1973), Patterson and Thompson (1975), Henderson (1977), and Miller (1977, 1979) and others. A brief description of these methods follows, based on Harville's (1977) review.

The log likelihood of the unknown parameters (β,θ) is proportional to

$$\ell(\underline{\beta},\theta,\underline{Y}) = -\log|V| - (\underline{Y}-X\underline{\beta})'V^{-1}(\underline{Y}-X\underline{\beta}) \tag{1.3.3}$$

The proper ML estimators for (β,θ) are a vector of values $(\underline{\hat{\beta}},\underline{\hat{\theta}})$ such that

$$\ell(\underline{\hat{\beta}},\hat{\theta},\underline{Y}) = \sup_{\beta,\theta\epsilon F} \ell(\underline{\beta},\theta,\underline{Y}) \tag{1.3.4}$$

It is reported that such an estimator does not exist in an important case considered by Focke and Dewess (1972). Neyman and Scott (1948) pointed out that ML estimators of variance components are heavily biased and in some cases they are not even consistent. In such cases, the use of ML estimators for drawing inferences on individual parameters may lead to gross errors, unless the exact distribution of the ML estimators is known. The drawbacks and the computational difficulties involved in obtaining the ML estimators palce some limitations on the use of the ML method in practical problems.

1.3.2 Maximum Likelihood Equations. We assume that $V > 0$(i.e., p.d.) for $\theta \in F$. Taking the derivatives of (1.3.3) w.r.t. β and θ_i and equating them to zero, we obtain the ML equations.

$$X'V^{-1}X\underline{\beta} = X'V^{-1}\underline{Y} \tag{1.3.5}$$

$$\operatorname{tr} V^{-1}v_i = (\underline{Y}-X\underline{\beta})'V^{-1}v_iV^{-1}(\underline{Y}-X\underline{\beta}),\ i=1,2,\ldots,k \tag{1.3.6}$$

Substituting for $\underline{\beta}$ in (1.3.6) from (1.3.5), the equations become

$$X\underline{\beta} = P\underline{Y},\ P = X(X'V^{-1}X)^{-}X'V^{-1}, \tag{1.3.7}$$

$$[T(\theta)]\theta = t_I(Y,\theta) \tag{1.3.8}$$

where $T(\theta) = (\operatorname{tr} V^{-1}v_iV^{-1}v_j)$ is a matrix and the i^{th} element $t_I(Y,\theta)$ is

$$\underline{Y}'(I-P)'V^{-1}v_iV^{-1}(I-P)\underline{Y}. \tag{1.3.9}$$

We have the following comments about the equations (1.3.7) and (1.3.8).

(i) The original ML equation (1.3.6) is unbiased while (1.3.8) which provides a direct estimate of θ is not so in the sense

$$E[t_I(Y,\theta)] \neq [T(\theta)]\theta. \tag{1.3.10}$$

An alternative to equation (1.3.8) is the one obtained by equating $t_I(Y,\theta)$ to its expectation

$$t_I(Y,\theta) = E[t_I(Y,\theta)] = [T_{UI}(\theta)]\theta \tag{1.3.11}$$

which is the marginal ML equation suggested by Patterson and Thompson (1975).

(ii) There may be no solution to (1.3.8) in the admissible set F to which θ belongs. This may happen when the supremum of the likelihood is attained at a boundary point of F.

(iii) The ML equation (1.3.8) is the same as that suggested by C. R. Rao (1979) for iterated MINQUE with the invariance property.

(iv) Maximum likelihood estimator of θ is invariant for translation of Y by $X\alpha$ for α (where α is an apriori value of θ).

(v) Computational algorithms: The equation (1.3.8) for the estimation of θ is, in general, very complicated and no closed form solution is possible. One has to adopt iterative procedures. Harville (1977) reviewed some of the existing methods:

(a) If $\hat{\theta}_K$ is the n^{th} approximation to the solution of (1.3.8), then $(k+1)^{th}$ approximation is

$$\hat{\theta}_{k+1} = [T(\hat{\theta}_k)]^{-1} t_I(Y,\hat{\theta}_k) \tag{1.3.12}$$

Equation (1.3.12) is suggested by Rao (1979) for iterative MINQUE with invariance, provided by θ is identifiable. Otherwise, the T matrix in (1.3.8) is not invertable. Iterative procedures of the type (1.3.12) are mentioned by Anderson (1973), LaMotte (1973) and Harville (1977) in different contexts. However, it is not known procedure (1.3.12) converges or whether it provides a solution at which the supremum of the likelihood is attained.

(b) Hemmerle and Hartley (1973) and Godnight and Hennerle (1978) suggested the method of W transformation for solving the ML equations. Miller (1979) has proposed a different approach. Harville (1977) also mentioned the variable-metric algorithms of Davidson-Fletcher-Powell described by Powell (1970). Further investigation is necessary for finding a satisfactory method of solving equation (1.3.8) and ensuring that the solution maximizes to likelihood.

1.3.2 Marginal Maximum Likelihood Equations. We already observed that the ML equation derived from (1.3.8) is not unbiased, since

$$E[t_I(Y,\theta)] \neq [T(\theta)]\theta. \qquad (1.3.13)$$

However, we may replace equation (1.3.8) by

$$t_I(Y,\theta) = E[t_I(Y,\theta)] = [T_{UI}(\theta)]\theta \qquad (1.3.14)$$

Equation (1.3.14) is obtained by Patterson and Thompson (1975) by maximizing the likelihood of θ based on L'Y, where L' is any choice of $X^{\perp}$, which is maximal invariant of $\underline{Y}$. Now

$$\ell(\theta,L',\underline{Y}) = -\log|L'VL| - \underline{Y}'L(L'VL)^{-1}L'\underline{Y} \qquad (1.3.15)$$

Differentiating (1.3.15) w.r.t. θ_i, we obtain the MML equation

$$tr[L(L'VL)^{-1}L'v_i] = \underline{Y}'L(L'VL)^{-1}L'v_iL(L'VL)^{-1}L'\underline{Y}, \qquad (1.3.16)$$

$$i=1,\ldots,k.$$

Using the identity (C. R. Rao, 1973, p. 77)

$$L(L'VL)^{-1}L' = V^{-1} - V^{-1}X(X'V^{-1}X)X'V^{-1} = V^{-1}(I-P) \qquad (1.3.17)$$

Equation (1.3.16) becomes

$$tr(V^{-1}(I-P)v_i) = \underline{Y}'V^{-1}(I-P)v_i(I-P')V^{-1}\underline{Y}, \quad i=1,2,\ldots,k \qquad (1.3.18)$$

which is independent of the choice of $L = X^{\perp}$ used in the construction of the maximal invariant of Y. It is easy to see that Equation (1.3.18) can be written as

$$[T_{UI}(\theta)]\theta = t_I(Y,\theta) \qquad (1.3.19)$$

which is Equation (1.3.14).

(i) Both the ML and MML estimates depend on the maximal invariant L'Y of Y. Neither method is applicable when θ is not identifiable on the basis of L'Y.

(ii) The bias in MMLE may not be as heavy as in MLE and the MMLE may be more useful as a point estimator.

(iii) The solution of (1.3.18) may not lie in the admissible set of θ.

(iv) If $\hat{\theta}_k$ is the k^{th} approximation, then the $(k+1)^{th}$ approximation can be obtained as

$$\hat{\theta}_{k+1} = [T_{UI}(\hat{\theta}_k)]^{-1}t_I(Y,\hat{\theta}_k) \qquad (1.3.20)$$

It is not known whether the process converges or yields a solution which maximizes the marginal likelihood.

It would seem that the computational effort required for each of MINQUE, MLE, and MMLE is much the same.

1.4 Least Square Method (LS Method)

The principle of LS, while conceptually quite distinct from the ML method and possessed of its own optimum properties, coincides with the ML method when the observations are normally distributed. It also provides unbiased estimators, linear in the observations, which have minimum variance. The assumption of normality is not required to establish this optimal property of LS estimators.

<u>1.4.1 Least Square Estimation in the Linear Model</u>. Consider the linear model

$$\underline{Y} = X\underset{\sim}{\beta} + \underline{\epsilon} \tag{1.4.1}$$

where $\underline{Y}$ is an $(n\times 1)$ vector of observations, X is an $(n\times k)$ matrix of known coefficient $(n>k)$, $\underset{\sim}{\beta}$ is a $(k\times 1)$ vector of parameters and $\underline{\epsilon}$ is an $(n\times 1)$ vector of random variables with $E(\underline{\epsilon}) = 0$ (1.4.2)

and dispersion matrix $V(\underline{\epsilon}) = E(\underline{\epsilon\epsilon'}) = \sigma^2\underline{I}$ (1.4.3)

The elements of $\underline{\epsilon}$ are uncorrelated and I is an $(n\times n)$ identity matrix. The LS method requires the minimization of the residual sum of squares.

$$S = (\underline{Y}-X\underset{\sim}{\beta})'(\underline{Y}-X\underset{\sim}{\beta}). \tag{1.4.4}$$

We obtain the LS estimation of $\underset{\sim}{\beta}$ by differentiating (1.4.4) and equating the derivative to zero, whence

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underline{Y} \tag{1.4.5}$$

Note that $\hat{\underset{\sim}{\beta}}$ is exactly the same as MLE of $\underset{\sim}{\beta}$.

We assume that $(X'X)$, the matrix of sum of squares and products of the elements of the column-vector composing X, is non-singular and can therefore be inverted.

1.4.2 Properties of LS Method. (i) Unbiasedness: Rewriting (1.4.5), we obtain

$$\hat{\beta} = (X'X)^{-1}X'(X\beta+\epsilon) = \beta + (X'X)^{-1}X'\epsilon.$$

Hence, using (1.4.2)

$$E(\hat{\beta}) = \beta \tag{1.4.6}$$

and $Var(\hat{\beta}) = E\{(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\} = E\{(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}\}$; using (1.4.3), we get

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}. \tag{1.4.7}$$

(ii) $\hat{\beta}$ is the LSE of β which minimized the residual sum of squares $\epsilon'\epsilon$, irrespective of any distribution properties of the error, but for testing hypothesis concerning the parameters, distributional assumptions are necessary.

(iii) The elements of $\hat{\beta}$ are linear functions of the observation Y, and provide unbiased estimates of the β which have minimum variances, irrespective of distribution properties of the error.

(iv) Unbiased estimation of the variance:

Consider the set of residuals in LS estimation

$$Y-X\hat{\beta} = (X\beta+\epsilon) - X\{(X'X)^{-1}X'(X\beta+\epsilon)\} . \tag{1.4.8}$$

After simplification, (1.4.8) becomes

$$Y-X\hat{\beta} = \{I_n-X(X'X)^{-1}X'\}\epsilon \quad . \tag{1.4.9}$$

The right-hand side of matrix in braces of (1.4.9) is symmetric and independent. So the sum of squares of residuals is

$$(\underset{\sim}{Y}-X\hat{\beta})'(\underset{\sim}{Y}-X\hat{\beta}) = \underline{\epsilon}'[I_n - X(X'X)^{-1}X']\underline{\epsilon} \tag{1.4.10}$$

$$\begin{aligned} E(\underset{\sim}{Y}-X\hat{\beta})'(\underset{\sim}{Y}-X\hat{\beta}) &= E(\underline{\epsilon}'M\underline{\epsilon});\text{ where } M = (I_n - X(X'X)^{-1}X') \\ &= tr[E(M\underline{\epsilon\epsilon}')] = \hat{\sigma}^2\, tr(M) \\ &= \hat{\sigma}^2\, tr[I_n - X(X'X)^{-1}X'] \\ &= \hat{\sigma}^2[tr\ I_n - tr\ I_k] \\ &= (n-k)\hat{\sigma}^2 . \end{aligned} \tag{1.4.11}$$

Thus an unbiased estimator of $\hat{\sigma}^2$ is, from (1.4.11),

$$s^2 = (\underset{\sim}{Y}-X\hat{\beta})'(\underset{\sim}{Y}-X\hat{\beta})/(n-k).$$

(v) If error vectors are independent and $\epsilon \sim N(0,\sigma^2 I)$, then $\hat{\underset{\sim}{\beta}}$ is the MLE of $\underset{\sim}{\beta}$.

1.5 Weighted Least Square (WLS) Method

The least square estimators might not be "best" when the components of the error vector $\underline{\epsilon}$ do not all "have the same chance of being small." The algebra of the Gauss-Markov theorem suggests the appropriate modification to the method of LS when the errors have different variances or when they are correlated.

1.5.1 WLS Estimation in the Linear Model.

Consider the model

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon} \tag{1.5.1}$$

where $E(\underline{\epsilon}) = 0$ and $V(\underline{\epsilon}) = \sigma^2 V$, where $V(\neq I)$ is a known positive definite matrix.

Let

$$P'P = PP = P^2 = V \tag{1.5.2}$$

where P is a non-singular matrix and set

$$Z = P^{-1}\underline{Y},\ A = P^{-1}X,\ \Phi = P^{-1}\underline{\epsilon}\ . \qquad (1.5.3)$$

Premultiplying (1.5.1) by P^{-1}, we obtain a new model as

$$P^{-1}\underline{Y} = P^{-1}X\underline{\beta} + P^{-1}\underline{\epsilon} \qquad (1.5.4)$$

applying (1.5.3) in (1.5.4), we get the model as

$$Z = A\underline{\beta} + \Phi \qquad (1.5.5)$$

and $E(\Phi) = 0$, $Var(\Phi) = \sigma^2 I$.

Using the Gauss-Markov theorem on (1.5.5), the estimator for β is

$$\hat{\underline{\beta}} = (A'A)^{-1}A'Z = [(X'(P^{-1})'P^{-1}X)^{-1}X'(P^{-1})'P^{-1}\underline{Y}]$$

or

$$\hat{\underline{\beta}} = (X'V^{-1}X)^{-1}X'V^{-1}Y. \qquad (1.5.6)$$

Similarly, the covariance matrix of $\hat{\underline{\beta}}$ is

$$V(\hat{\underline{\beta}}) = (A'A)^{-1}\sigma^2 = [X'(P^{-1})'P^{-1}X]^{-1}\sigma^2 = (X'V^{-1}X)^{-1}\sigma^2. \qquad (1.5.7)$$

In practice, the WLS method is mostly applied when the observations are independent but have different variances. It may be difficult to obtain specific information on the form of V initially. For this reason, it is sometimes necessary to begin with the assumption $V = \sigma^2 I$ and then attempt to discover something about the form of V by examining the residuals from the regression analysis.

If a WLS analysis was called for but a LS analysis was performed, the estimates would still be unbiased but would not have minimum variance.

The WLS method was first developed by Aitken (1935). Goldman and Zelen (1964) were the first to consider the case when V is singular.

1.6 Purpose of the Thesis

It is the purpose of this thesis to develop the theoretical properties and to compare the performance of the different estimators (e.g., OLS, WLS, ML, MINQUE, Modified MINQUE, Posterior Likelihood (PL), etc.) for the linear model when variances are unknown and different. We also wish to check the relative performance of large smaple properties and small sample results for both normal and non-normal distributions.

We have already presented a brief review of the Ordinary Least Squares (OLS) estimators and its modified form Weighted Least Squares (WLS) estimators with their properties in Sections 1.4 and 1.5 respectively. Another approach is estimation by maximum likelihood assuming that the observations are normally distributed, which is given in Section 1.3. Another approach to the above problem is estimation by quadratic functions of observations, based on sums of squares appearing in the analysis of variance table (e.g., Henderson, 1953; Searle, 1968, 1971). As remarked by Rao (1972), "In this method the theoretical basis is not clear, the procedures suggested are ad hoc and much seems to depend on intuition." This led C. R. Rao (1970) to introduce the principle of MINQUE. In his paper, he suggested using the estimated variances obtained through MINQUE principle to carry out the WLS procedure for obtaining the estimates of regression in parameters.

There are some drawbacks in this method. The major drawback of MINQUE is that it may give negative values for estimates of non-negative variances, as noted by many authors (see Section 1.2). J. N. K. Rao (1973) gave some modifications of MINQUE based on intuitive grounds. A complete and detailed review of MINQUE technique with its modifications was presented in Section 1.2.

Due to the major drawbacks of the MINQUE technique, we also develop another procedure on the basis of prior likelihoods called Posterior Likelihood (PL) estimation, which is discussed and compared with other techniques, in Chapter 4.

1.7 Outline of the Thesis

Chapter 1 provides the review of the literature on the existing techniques for estimation of variances in the linear model.

In Chapter 2, we develop and present several theorems on the large sample variances of estimates for the regression parameters for both WLS and MINQU-based estimators for normal and non-normal distributions. The main purpose of developing these theorems is to provide a theoretical basis for comparing the large sample properties of WLS and MINQU-based estimators with small samples simulated results.

Chapter 3 provides a comparison of the theoretical results based on Chapter 2 with simulated small samples results. These simulated and theoretical results are based on WLS and MINQU-based estimators for normal and non-normal distributions. A Monte Carlo comparison of the modified MINQU-based, OLS and WLS estimators is also given.

In Chapter 4, we present the theoretical developments of the Posterior Likelihood (PL) methodology and their theoretical properties. Some theoretical results about the $\underline{\hat{\beta}}_1$ and the variances are obtained using Gamma prior likelihoods for the regression model. A comparative empirical study on the basis of OLS, WLS, ML, and PL is also presented.

In Chapter 5, we present adaptive procedures for estimation. A review of existing techniques for testing equality of variances is presented. Adaptive estimation procedures for the estimation of regression parameters after using preliminary tests of variance equality are discussed and empirical results on the adaptive estimators are also provided.

Chapter 6 provides a review of the literature on multiple comparisons and adaptive estimators using multiple comparisons procedures. Empirical comparison of these adaptive estimators with other estimators are also presented.

In Chapter 7, we examine the effects of "errors in variables" upon the different estimators. Empirical comparisons are again provided.

Finally, in Chapter 8, we present an overview of the results and summarize the main findings of the study.

CHAPTER 2

LARGE SAMPLE PROPERTIES OF WLS AND MINQU-BASED ESTIMATORS

2.1 Introduction

The main purpose of this chapter is to determine the large sample properties of WLS and MINQU-based estimators for the linear model. The theoretical results of this chapter will be used in the next chapter for comparison with the simulated small samples results.

In Section 2.2, we state a general Theorem 2.2.1 for the WLS and MINQUE cases with normally distributed error terms, with zero means and variances σ_{ii} for the n_i observations corresponding to X_i, $i=1,\ldots,k$. We also develop several lemmas prior to prove the main theorem.

In Section 2.3, we present the complete proofs of Theorem 2.2.1 for the WLS and MINQUE cases.

In Section 2.4, we present and prove Theorem 2.4.1 for WLS and MINQU-based estimators for non-normal distributions. Two new lemmas are also derived to prove the results of Theorem 2.4.1.

To check the applicability of the above theorems for WLS and MINQU-based estimators for normal and non-normal cases, we consider some special cases of the above theorems in Section 2.5. In particular, we consider the special cases when all the variances are equal and all the $n_i = \frac{n}{k}$. Results are given for each of WLS and MINQUE for both normal and non-normal distributions.

2.2 Asymptotic Results

In the next two sections, we shall state and prove theorems which give the expected value and variance of WLS and MINQU-based estimators for $\underline{\beta}$ in the linear model.

Theorem 2.2.1

Consider the linear model

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon}, \tag{2.2.1}$$

where $\underline{Y}$ is a vector of $n = \sum_{i=1}^{k} n_i$ observations, X is a known matrix of order $n \times p$, $\underline{\beta}$ is a vector of p unknown parameters and $\underline{\epsilon}$ is a vector of n. Further, assume that

$$\underline{\epsilon}_i \sim \text{NID}(\underline{0}, \sigma_{ii} I_{n_i}). \tag{2.2.2}$$

(i) The WLS Case

The WLS estimator is

$$\hat{\underline{\beta}}_W = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{Y}, \tag{2.2.3}$$

where

$$\hat{V} = \begin{bmatrix} \hat{\sigma}_{11} I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{kk} I_{n_k} \end{bmatrix}; \tag{2.2.4}$$

and $\hat{\sigma}_{ii} = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$; $i=1,2,\ldots,k$; $j=1,2,\ldots,n_i$.

Then $E(\hat{\underline{\beta}}_W) = \underline{\beta}_W$; and to terms of $0(n^{-2})$,

$$\text{Var}(\hat{\underline{\beta}}_W) = A_o^{-1} + 2A_o^{-1} x \,\text{diag}[\sigma_{ii}^{-2}(\sigma_{ii} - n_i g_i)]\, x' A_o^{-1} + 0(n^{-3}), \tag{2.2.5}$$

where

$$\underset{(p\times p)}{A_o} = \underset{(p\times k)}{x} \underset{(k\times k)}{\begin{bmatrix} n_1\sigma_{11}^{-1} & & 0 \\ & \ddots & \\ 0 & & n_k\sigma_{kk}^{-1} \end{bmatrix}} \underset{(k\times p)}{x'} \tag{2.2.6}$$

and x has rows $\underline{x}_i$, $(i=1,\ldots,k)$.

(ii) The MINQUE Case

The MINQU-based estimator for $\underline{\beta}$ is

$$\hat{\underline{\beta}}_M = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{Y}, \tag{2.2.7}$$

where
$$\hat{V} = \begin{bmatrix} \hat{\sigma}_{11}I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{kk}I_{n_k} \end{bmatrix}$$

and the MINQUE of $\hat{\sigma}_{ii} = \dfrac{n}{n_i(n-2)} \sum_i (y_{ij} - \bar{y})^2 - \dfrac{S^2}{n-2}$, (2.2.8)

where $S^2 = (n-1)^{-1}\Sigma\Sigma(y_{ij} - \bar{y})^2$.

Then $E(\hat{\underline{\beta}}_M) = \hat{\underline{\beta}}_M$, and to terms of $O(n^{-2})$,

$$\mathrm{Var}(\hat{\underline{\beta}}_M) = A_o^{-1} + 2A_o^{-1}x\ \mathrm{diag}[n_i t_{ii}/\sigma_{ii}{}^3]x'A_o^{-1} - 2A_o^{-1}x[n_i n_j t_{ij} g_{ij}/$$

$$\sigma_{ii}{}^2\sigma_{jj}{}^2]x'A_o^{-1} + O(n^{-3}). \tag{2.2.9}$$

where $t_{ij} = \mathrm{cov}(\hat{\sigma}_{ii}, \hat{\sigma}_{jj})$ and $g_{ij} = \underline{x}_i^1\ A_0^{-1}\underline{x}_j$.

In order to prove the results of the above Theorem 2.2.1, we first establish several lemmas.

Lemma 2.2.1

If the distribution of $\underline{\epsilon}$ is symmetric about zero, and $g(\underline{\epsilon})$ is an even function of $\underline{\epsilon}$ with finite expectation then $E\{\underline{\epsilon}g(\underline{\epsilon})\} = \underline{0}$, provided the expectation exists.

Proof:

Since $P(\underline{\epsilon}_i > 0) = \frac{1}{2}$

$$E[\epsilon_i g(\underline{\epsilon})] = E[\epsilon_i g(\underline{\epsilon})\,|\,\epsilon_i > 0]\tfrac{1}{2} + E[\epsilon_i g(\underline{\epsilon})\,|\,\epsilon_i < 0]\tfrac{1}{2}\ .$$

Let $\underline{\eta} = -\ \underline{\epsilon}$.

Then $E[\epsilon_i g(\underline{\epsilon})] = E[\epsilon_i g(\underline{\epsilon})\,|\,\epsilon_i > 0]\frac{1}{2} - E[\eta_i g(\underline{\eta})\,|\,\eta_i > 0]\frac{1}{2} = 0$

$$\therefore E[\underline{\epsilon}g(\underline{\epsilon})] = \underline{0} \tag{2.2.10}$$

Lemma 2.2.2

$$\hat{V}^{-1} = V^{-1} - V^{-1}(\hat{V}-V)V^{-1} + V^{-1}(\hat{V}-V)V^{-1}(\hat{V}-V)V^{-1}\ O\big((V-\hat{V})^3\big).$$

where $\hat{V}$ is any diagonal matrix.

Proof of Lemma 2.2.2

Directly:-

$$\hat{V}^{-1}-V^{-1} = -V^{-1}(\hat{V}-V)\hat{V}^{-1} \qquad (*)$$

Also $\quad \hat{V}^{-1} = (\hat{V}^{-1}-V^{-1}) + V^{-1} \qquad (**)$

so

$$\hat{V}^{-1} - V^{-1} = -V^{-1}(\hat{V}-V)V^{-1} - V^{-1}(\hat{V}-V)(V^{-1}-V^{-1})$$

and using * this $= -V^{-1}(\hat{V}-V)V^{-1} + V^{-1}(\hat{V}-V)V^{-1}(\hat{V}-V)\hat{V}^{-1}$.

The result follows directly by using ** then * recursively.

Lemma 2.2.3

When V is estimatee by WLS,

$$E(\hat{V}U\hat{V}) = VUV + 2 \text{ diag}\left(\frac{\sigma_{ii}{}^{2}u_{ii}}{n_i}\right)$$

where U is a known n n matrix with sub-matrices u_{ij} of order $n_i \times n_j$; and $\hat{V}$ is given in (2.2.4).

Proof:

$\hat{V}U\hat{V}$ has $(i,j)^{th}$ element is $= \hat{\sigma}_{ii}\ \hat{\sigma}_{jj}\ u_{ij}$ and $\hat{V} = \text{diag}(\hat{\sigma}_{ii})$;

$E(\hat{\sigma}_{ii}) = \sigma_{ii}$, i=j; $\hat{\sigma}_{ij} \equiv 0$, i≠j.

$$E\left[\hat{\sigma}_{ii}\ \hat{\sigma}_{jj}\ u_{ij}\right] = u_{ij}\left[\mathrm{cov}(\hat{\sigma}_{ii},\ \hat{\sigma}_{jj}) + E[\hat{\sigma}_{ii}]\ E[\hat{\sigma}_{jj}]\right]$$

$$E\left[\hat{\sigma}_{ii}\ \ {}_{jj}\ u_{ij}\right] = \begin{cases} u_{ij}\ \left[0 + \sigma_{ii}\ \sigma_{jj}\right], & i \neq j, \\ u_{ii}\ \left[\mathrm{var}(\hat{\sigma}_{ii}) + \sigma_{ii}^2\right], & i = j, \end{cases}$$

and the result follows since

$$\mathrm{var}(\hat{\sigma}_{ii}) = 2\sigma_{ii}^2/n_i.$$

<u>Lemma 2.2.4</u>

When V is estimated by MINQUE,

$$E(\hat{V}U\hat{V}) = VUV + B_1 + 2B_2$$

where $B_1 = [u_{ij}C_{ij}]$, $B_2 = \mathrm{diag}[u_{ii}w_{ii}]$, $w_{ii} = \dfrac{n\sigma_{ii}}{n_i(n-2)^2}\left[(n-4)\sigma_{ii}+2\bar{\sigma}\right]$

and $\quad \bar{\sigma} = \sum_{i=1}^{k} n_i\sigma_{ii}/n.$

U is a known matrix of $n \times n$ with sub-matrices u_{ij} of order $n_i \times n_j$ and

$$\hat{V} = \begin{bmatrix} \hat{\sigma}_{11} I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{kk} I_{n_k} \end{bmatrix} .$$

The MINQUE of $\hat{\sigma}_{ii}$ is given in (2.2.8).

<u>Proof</u>:

$E(\hat{V}U\hat{V})$ has the diagonal elements $E(u_{ii}\hat{\sigma}^2_{ii})$ and off-diagonal elements

$$E(u_{ij}\hat{\sigma}_{ii}\hat{\sigma}_{jj}) ;$$

$$E(\hat{\sigma}_{ii}) = \sigma_{ii} \quad \text{since MINQUE's are unbiased.}$$

Routine calculations show that

$$t_{ii} = \mathrm{Var}(\hat{\sigma}_{ii}) = 2(n-2)^{-2}\left[\sigma^2_{ii} \left\{ \frac{n}{n_i}(n-4) + \frac{2(n+1)}{(n-1)} \right\} + 2n\sigma_{ii}\bar{\sigma} \left\{ \frac{1}{n_i} - \frac{2}{n-1} \right\} + (n-1)^{-2}(n^2\bar{\sigma}^2 - \Sigma n_i \sigma^2_{ii}) \right] \qquad (2.2.15)$$

and

$$t_{ij} = C_{ij} = \mathrm{Cov}(\hat{\sigma}_{ii},\hat{\sigma}_{jj}) = 2(n-2)^{-2} \left[2\left(\sigma_{ii}\sigma_{jj} + \frac{\sigma^2_{ii}+\sigma^2_{jj}}{n-1} \right) - \frac{2n}{n-1}(\sigma_{ii}+\sigma_{jj})\bar{\sigma} + (n-1)^{-2}(n^2\bar{\sigma}^2 - \Sigma n_i \sigma^2_{ii}) \right] ; \quad i \neq j \qquad (2.2.16)$$

so that

$$t_{ii} = C_{ii} + \frac{2n\sigma_{ii}}{n_i(n-2)^2}\left[(n-4)\sigma_{ii}+2\bar{\sigma}\right]$$

$$= C_{ii} + 2W_{ii}.$$

As $E(u_{ii}\hat{\sigma}_{ii}^2) = u_{ii}\sigma_{ii}^2 + u_{ii}\ \mathrm{Var}(\hat{\sigma}_{ii})$

and $E(u_{ij}\hat{\sigma}_{ii}\hat{\sigma}_{jj}) = u_{ij}\sigma_{ii}\sigma_{jj} + u_{ij}\ \mathrm{Cov}(\hat{\sigma}_{ii},\hat{\sigma}_{jj})$

it follows that

$$E(\hat{V}U\hat{V}) = VUV + B_1 + 2B_2. \tag{2.2.17}$$

Lemma 2.2.5

Consider terms of the form

$$U = HT(\underline{\epsilon\epsilon}'-V)$$

where $T_{(n\times n)} = \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1k} \\ T_{21} & T_{22} & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ T_{k1} & \cdots & \cdots & T_{kk} \end{pmatrix}$

is a matrix of known terms, T_{ij} is $(n_i \times n_j)$ and $n = \sum_{i=1}^{k} n_i$

$H = \hat{V}-V$, $\quad V = \begin{pmatrix} \sigma_{11}I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \sigma_{kk}I_{n_k} \end{pmatrix}$; $\hat{V}$ has $\hat{\sigma}_{ii}$ in place of σ_{ii}

and $$\underset{n\times 1}{\epsilon} = \begin{bmatrix} \underline{\epsilon}_1 \\ \vdots \\ \underline{\epsilon}_k \end{bmatrix} \quad \text{where } \underline{\epsilon}_i = \begin{bmatrix} \epsilon_{i,1} \\ \vdots \\ \epsilon_{i,n_i} \end{bmatrix}.$$

Also $\hat{V}$ is the WLS estimator of V.

Then $E(U) = 2VDT_\Delta V$

where

T_Δ is a block diagonal matrix, $$T_\Delta = \begin{bmatrix} T_{11}M_{11} & & 0 \\ & \ddots & \\ 0 & & T_{kk}M_{kk} \end{bmatrix},$$

$$M_{rr} = (I - \frac{1}{n_r} \underline{1}\,\underline{1}'), \quad \text{and} \quad D = \begin{bmatrix} \frac{1}{n_1} I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{n_k} I_{n_k} \end{bmatrix}.$$

Proof:

U has $(r,q)^{th}$ block $$u_{rq} = \sum_s (\hat{\sigma}_{rr} - \sigma_{rr}) T_{rs} (\underline{\epsilon}_s \underline{\epsilon}_q' - \sigma_{sq} I_{sq})$$

$$= \sum_s U_{rsq} \text{ (say)}$$

where
$$\sigma_{sq} = \begin{cases} \sigma_{qq} & \text{if } s=q \\ 0 & \text{elsewhere} \end{cases}$$

$$I_{sq} = \begin{cases} 0 & \text{if } s \neq q \\ I_{qq} & \text{if } s=q \end{cases}$$

$$\text{Now } (n_r-1)\hat{\sigma}_{rr} = \sum_{i=1}^{n_r} (y_{r_i}-\bar{y}_r)^2$$

$$= \sum_{i=1}^{n_r} (\epsilon_{r_i}-\bar{\epsilon}_r)^2 \quad \text{since } y_{r_i} = x_r'\beta + \epsilon_{r_i}$$

$$\text{thus } E(U_{rsq}) = E\left[\frac{\left(\sum_{i=1}^{n_r} \epsilon_{r_i}^2 - n_r(\bar{\epsilon}_r)^2\right) T_{rs}\underline{\epsilon}_s\underline{\epsilon}_q'}{(n_r-1)}\right] - \sigma_{rr}T_{rs}\sigma_{sq}I_{sq}.$$

$$\left.\begin{array}{l} \text{If } r \neq s \neq q \\ \text{or } r = s,\ r \neq q \\ \text{or } r \neq s,\ r = q \end{array}\right\} \Longrightarrow E(U_{rsq}) = 0 \text{ directly.}$$

$$\text{If } s = q \neq r,\ E(U_{rqq}) = \sigma_{rr}\sigma_{qq}T_{rq} - \sigma_{rr}\sigma_{qq}T_{rq} = 0$$

$$\text{If } s = q = r,\ E(U_{rrr}) = \frac{1}{(n_r-1)} E\left[\left(\Sigma\epsilon_{r_i}^2 - n_r(\bar{\epsilon}_r)^2\right)T_{rr}\,\underline{\epsilon}_r\underline{\epsilon}_r'\right] - \sigma_{rr}^2 T_{rr} \tag{2.2.18}$$

$$\text{Consider } \underline{\epsilon}_r\underline{\epsilon}_r'\left(\Sigma\epsilon_{r_i}^2 - n_r(\bar{\epsilon}_r)^2\right) = \left[\epsilon_{ra}\epsilon_{rb}\left(\Sigma\epsilon_{r_i}^2 - n_r\bar{\epsilon}_r^2\right) \text{ of } (a,b)^{th} \text{ elements}\right]$$

$$\text{if } a \neq b,\ E\left[\underline{\epsilon}_r\underline{\epsilon}_r'\left(\Sigma\epsilon_{r_i}^2 - n_r(\bar{\epsilon}_r)^2\right)\right] = -\frac{2\sigma_{rr}^2}{n_r}$$

$$\text{if } a = b,\ E\left[\underline{\epsilon}_r\underline{\epsilon}_r'\left(\Sigma\epsilon_{r_i}^2 - n_r(\epsilon_r')^2\right)\right] = \sigma_{rr}^2\left[(n_r-1)+3\right] - \frac{\sigma_{rr}^2}{n_r}\left[(n_r-1)+3\right]$$

$$= \frac{(n_r+2)(n_r-1)}{n_r}\sigma_{rr}^2$$

Substituting these values (2.2.18) will be as

$$E(u_{rrr}) = \frac{2\sigma_{rr}^2}{n_r-1} T_{rr} \left[\underline{I}_{n_r} - \frac{1}{n_r} \underline{1}\,\underline{1}'\right]$$

$$E(u_{rq}) = 0, \text{ when } r \neq q.$$

$$E(U_{rq}) = \frac{2\sigma_{rr}^2 T_{rr}}{n_r-1} \left[\underline{I}_{n_r} - \frac{1}{n_r} \underline{1}\,\underline{1}'\right] ; \quad \text{when } r = q$$

$$= \frac{2\sigma_{rr}^2 T_{rr}}{n_r-1} M_{rr}.$$

$$\therefore E(U) = 2VDT_{\Delta}V. \tag{2.2.19}$$

Lemma 2.2.6

Consider terms of the form

$$U = HT(\underline{\epsilon\epsilon}'-V)$$

where
$$T = \begin{pmatrix} T_{11} & T_{12} & \cdots & \cdots \\ T_{21} & T_{22} & \cdots & \vdots \\ \cdots & \cdots & \cdots & T_{kk} \end{pmatrix}$$

is a matrix of known terms, T_{ij} is $(n_i \times n_j)$ and $n = \sum_{i=1}^{k} n_i$

$H = \hat{V}-V$, where $V = \begin{pmatrix} \sigma_{11}I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \sigma_{kk}I_{n_k} \end{pmatrix}$ and $\hat{V}$ has $\hat{\sigma}_{ii}$ in place of σ_{ii}.

The MINQUE of $\hat{\sigma}_{ii}(M)$ is given in (2.2.8) and

$$\underset{n\times 1}{\underline{\epsilon}} = \begin{pmatrix} \underline{\epsilon}_1 \\ \vdots \\ \underline{\epsilon}_k \end{pmatrix} \quad \text{where } \underline{\epsilon}_i = \begin{pmatrix} \epsilon_{i,1} \\ \vdots \\ \epsilon_{i,n_i} \end{pmatrix} .$$

Then $E(U) = 2CVDT_\Delta V$

$$\text{where} \qquad C = \text{diag}\left[\left\{\frac{n}{n_r(n-2)} - \frac{(n_r-1)}{(n-1)(n-2)}\right\} I_{n_r}\right] ,$$

$$T_\Delta \text{ is the block diag matrix, } T_\Delta = \begin{pmatrix} T_{11}M_{11} & & 0 \\ & \ddots & \\ 0 & & T_{kk}M_{kk} \end{pmatrix} ,$$

$$M_{rr} = (I - \frac{1}{n_r}\underline{1}\,\underline{1}') \quad \text{and} \quad D = \begin{pmatrix} \frac{1}{n_1} I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{n_k} I_{n_k} \end{pmatrix}$$

Proof:

U has $(r,q)^{th}$ block $u_{rq} = \sum_s (\hat{\sigma}_{rr}(M) - \sigma_{rr}(M))T_{rs}(\underline{\epsilon}_s\underline{\epsilon}_q' - \sigma_{sq}I_{sq})$

$$\text{where} \quad \hat{\sigma}_{rr}(M) = \frac{n\hat{\sigma}_{rr}}{n_r(n-2)} - \frac{\sum_r (n_r-1)\hat{\sigma}_{rr}}{(n-1)(n-2)} .$$

From Lemma 2.2.5

$$E(u_{rrr}(M)) = \frac{2\sigma_{rr}^2}{n_r-1} T_{rr} M_{rr} \left[\frac{n}{n_r(n-2)} - \frac{n_r-1}{(n-1)(n-2)} \right]$$

$$E(U) = 2VCDT_\Delta V \ . \tag{2.2.20}$$

Lemma 2.2.7

If X is a $n \times p$ matrix $\left(\sum_{i=1}^{k} n_i = n \right)$ and $A_o = X'V^{-1}X$.

Then $\mathrm{diag}(XA_o^{-1}X') = \begin{pmatrix} g_1 I_{n_1} & & 0 \\ & \ddots & \\ 0 & & g_k I_{n_k} \end{pmatrix}$

where $g_i = \underline{x}_i' A_o^{-1} \underline{x}_i$, where the $n_{i-1}+1, n_{i-1}-2, \ldots, n_i{}^{th}$ rows of X are $\underline{x}_i'$.

Proof:

Since $\underset{(n\times p)}{X} = \begin{pmatrix} \underline{x}_1' \\ \vdots \\ \underline{x}_1' \\ \hline \underline{x}_2' \\ \vdots \\ \hline \vdots \\ \hline \underline{x}_k' \\ \vdots \\ \underline{x}_k' \end{pmatrix}$

We can write matrix X as

$$X = \begin{pmatrix} \underset{(n_1\times 1)}{\underline{1}_1} & & & 0 \\ & \underset{(n_2\times 1)}{\underline{1}_2} & & \\ & & \ddots & \\ 0 & & & \underset{(n_k\times 1)}{\underline{1}_k} \end{pmatrix} \begin{pmatrix} \underline{x}_1' \\ \vdots \\ \underline{x}_k' \end{pmatrix} = \underset{(n\times k)}{L} \; \underset{(k\times p)}{x'} \tag{2.2.21}$$

where

$$\underset{(k\times p)}{x'} = \begin{pmatrix} \underline{x}_1' \\ \vdots \\ \underline{x}_k' \end{pmatrix} \quad \text{and} \quad \underline{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} .$$

Thus $XA_o^{-1}X' = Lx'A_o^{-1}xL'$.

$$\text{Hence} \quad \operatorname{diag}(XA_o^{-1}X') = \begin{pmatrix} \underline{x}_1'A_o^{-1}\underline{x}_1 I_{n_1} & & & \\ & \underline{x}_2'A_o^{-1}\underline{x}_2 I_{n_2} & & 0 \\ & & \ddots & \\ & 0 & & \underline{x}_k'A_o^{-1}\underline{x}_k I_{n_k} \end{pmatrix}$$

$$= \begin{pmatrix} g_1 I_{n_1} & & \\ & g_2 I_{n_2} & 0 \\ & 0 \quad \ddots & \\ & & g_k I_{n_k} \end{pmatrix} \text{, by definition.} \tag{2.2.22}$$

Lemma 2.2.8

If M is the diagonal matrix

$$M = \begin{pmatrix} M_1 & & 0 \\ & \ddots & \\ 0 & & M_k \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} \underset{n_1\times 1}{\underline{1}_1} & & & 0 \\ & \underset{n_2\times 1}{\underline{1}_2} & & \\ & & \ddots & \\ 0 & & & \underset{n_k\times 1}{\underline{1}_k} \end{pmatrix}.$$

where $M_{rr} = (I_{n_r} - \frac{1}{n_r}\underline{1}_r\underline{1}_r')$, then $ML = 0$.

Proof:

In matrix ML, all the terms on the block diagonal are alike, so it will be sufficient to show that

$$M_{rr}\underline{1}_r = \underline{0}$$

$$ML = \begin{pmatrix} M_1\underline{1}_1 & & 0 \\ & \ddots & \\ 0 & & M_k\underline{1}_k \end{pmatrix} \quad ; \text{ where } M_{rr}\underline{1}_r = (I_{n_r} - \frac{1}{n_r}\underline{1}_r\underline{1}_r')\underline{1}_r = \underline{1}_r - \underline{1}_r = \underline{0}. \tag{2.2.23}$$

Hence $ML = 0$.

Lemma 2.2.9

$$X'V^{-1}HV^{-1}X = xh_\Delta x'$$

where $$\underset{(k\times k)}{h_\Delta} = \operatorname{diag}\left[\frac{n_i(\hat{\sigma}_{ii}-\sigma_{ii})}{\sigma_{ii}^2}\right].$$

Proof:

From (2.2.21), $X = Lx'$

hence $X'V^{-1}HV^{-1}X = xLV^{-1}HV^{-1}Lx'$ (2.2.24)

But $$V^{-1}HV^{-1} = \mathrm{diag}\left[\left(\frac{\hat{\sigma}_{ii}-\sigma_{ii}}{\sigma_{ii}^2}\right) I_{n_i}\right]$$

so that (2.2.24) reduces, as required, to

$$X'V^{-1}HV^{-1}X = xh_{\Delta}x'. \tag{2.2.25}$$

2.3 Proof of Theorem 2.2.1

(i) The WLS Case

The linear model is $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$.

The WLS estimator is

$$\hat{\underline{\beta}}_W = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{Y}$$

where $\hat{V} = \begin{bmatrix} \hat{\sigma}_{11}I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{kk}I_{n_k} \end{bmatrix}$ and $\hat{\sigma}_{ii} = \sum_{j=1}^{n_i} (y_{ij}-\bar{y}_i)^2/(n_i-1)$.

It follows that

$$\hat{\underline{\beta}}_W = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}(X\underline{\beta}+\underline{\epsilon}) = \underline{\beta} + (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{\epsilon}.$$

Hence, $E(\hat{\underline{\beta}}_W) = \underline{\beta} + E\ (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{\epsilon}$.

Since $\hat{V}$ is an even function of $\underline{\epsilon}_1,\ldots,\underline{\epsilon}_n$, it follows that $(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}$ is an even function of $\underline{\epsilon}_1,\ldots,\underline{\epsilon}_n$. Thus, applying Lemma 2.2.1,

$$E(\hat{\underline{\beta}}_W) = \underline{\beta}_W.$$

Now

$$\mathrm{Var}(\hat{\underline{\beta}}_W) = E(\hat{\underline{\beta}}_W-\underline{\beta})(\hat{\underline{\beta}}_W-\underline{\beta})' = E\{(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{\epsilon\epsilon}'\hat{V}^{-1}X(X'\hat{V}^{-1}X)^{-1}\}. \tag{2.3.1}$$

$$\text{Let } \underline{\epsilon\epsilon}' = V + (\underline{\epsilon\epsilon}'-V). \tag{2.3.2}$$

$$\text{Since } E(\underline{\epsilon\epsilon}') = V = E(\hat{V}), \tag{2.3.3}$$

$$\mathrm{Var}(\hat{\underline{\beta}}_W) = E\ (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\{V+(\epsilon\epsilon'-V)\}\hat{V}^{-1}X(X'\hat{V}^{-1}X)^{-1} \tag{2.3.4}$$

Applying Lemma 2.2.2 and substituting the values of $\hat{V}^{-1}$ into the term $(X'\hat{V}^{-1}X)$,

$$\begin{aligned}(X'\hat{V}^{-1}X) &= X'\{V^{-1}-V^{-1}(\hat{V}-V)V^{-1}+V^{-1}(\hat{V}-V)V^{-1}(\hat{V}-V)V^{-1}\}X \ +\cdots \\ &= X'V^{-1}X-X'V^{-1}(\hat{V}-V)V^{-1}X+X'V^{-1}(\hat{V}-V)V^{-1}(\hat{V}-V)V^{-1}X \ +\cdots \ .\end{aligned}$$

Let $\hat{V} - V = H$,

$X'V^{-1}X = A_0$,

$X'V^{-1}(\hat{V}-V)V^{-1}X = X'V^{-1}HV^{-1}X = A_1$,

and $X'V^{-1}(\hat{V}-V)V^{-1}(\hat{V}-V)V^{-1}X = X'V^{-1}HV^{-1}HV^{-1}X = A_2$,

then

$$(X'\hat{V}^{-1}X) = A_0 - A_1 + A_2 + \text{terms of } O(H^3) \tag{2.3.5}$$

$$\begin{aligned}\therefore\ (X'\hat{V}^{-1}X)^{-1} &= A_0^{-1} + A_0^{-1}A_1A_0^{-1} - A_0^{-1}A_2A_0^{-1} + A_0^{-1}A_1A_0^{-1}A_1A_0^{-1} \\ &\quad + \text{terms of } O(H^3).\end{aligned} \tag{2.3.6}$$

Substituting the values of (2.3.6) up to $O(H^3)$ in (2.3.4)

$$\mathrm{Var}(\hat{\underline{\beta}}_W) = E[FX'\{V^{-1}-V^{-1}HV^{-1}+V^{-1}(HV^{-1})^2\}\{V+(\underline{\epsilon\epsilon}'-V)\}\{V^{-1}-V^{-1}HV^{-1} + V^{-1}(HV^{-1})^2\}XF] \qquad (2.3.7)$$

where $F = A_0^{-1} + A_0^{-1}A_1A_0^{-1} - A_0^{-1}A_2A_0^{-1} + A_0^{-1}(A_1A_0^{-1})^2$.

Thus

$$\mathrm{Var}(\hat{\underline{\beta}}_W) = E[FX'V^{-1}(I-HV^{-1})(\underline{\epsilon\epsilon}'-V)(I-HV^{-1})'V^{-1}XF] + E[FX'V^{-1}\{I-HV^{-1} + (HV^{-1})^2\}^2XF], \qquad (2.3.8)$$

which we write as

$$\mathrm{Var}(\hat{\underline{\beta}}_W) = \text{Term I} + \text{Term II}. \qquad (2.3.9)$$

We will now examine these two terms separately.

$$\begin{aligned}\text{Term II} &= E[FX'V^{-1}(I-HV^{-1}+(HV^{-1})^2)^2XF] \\ &= E[FX'V^{-1}\{I-2HV^{-1}+3(HV^{-1})^2+O(H^3)\}XF] \\ &= E[F\{X'V^{-1}X-2X'V^{-1}HV^{-1}X+3X'V^{-1}(HV^{-1})^2X\}F] \\ &= E[F(A_0^{-1}-2A_1+3A_2)F] \\ &= E[\{A_0^{-1}+A_0^{-1}A_1A_0^{-1}-A_0^{-1}A_2A_0^{-1}+A_0^{-1}(A_1A_0^{-1})^2\}(A_0-2A_1+3A_2)\{A_0^{-1} \\ &\quad + A_0^{-1}A_1A_0^{-1}-A_0^{-1}A_2A_0^{-1}+A_0^{-1}(A_1A_0^{-1})^2+O(H^3)\}] \\ &= E[A_0^{-1}+A_0^{-1}A_2A_0^{-1}-A_0^{-1}(A_1A_0^{-1})^{]}+O(n^{-3})]. \end{aligned} \qquad (2.3.10)$$

$$\text{Term I} = E[FX'V^{-1}(I-HV^{-1})(\underline{\epsilon\epsilon}'-V)(I-HV^{-1})'V^{-1}XF]$$

Let $\underline{\epsilon\epsilon}' - V = Q$, then

Then, Term I $= E[FX'V^{-1}(I-HV^{-1})Q(I-HV^{-1})'V^{-1}XF]$

$$= E[A_0^{-1}X'V^{-1}QV^{-1}XA_0^{-1}-A_0^{-1}X'V^{-1}HV^{-1}QV^{-1}XA_0^{-1}-A_0^{-1}X'V^{-1}QV^{-1}HV^{-1}XA_0^{-1}$$

$$+ A_0^{-1}A_1A_0^{-1}X'V^{-1}QV^{-1}XA_0^{-1}+A_0^{-1}X'V^{-1}QV^{-1}XA_0^{-1}A_1A_0^{-1}+O(n^{-3})]. \tag{2.3.11}$$

Then we must evaluate the following terms:

$E(A_2)$, $E(A_1A_0^{-1}A_1)$ in Term II

$E(QT_1A_1)$, $E(A_1T_1Q)$ in Term I.

$$E(A_2) = E(X'V^{-1}HV^{-1}HV^{-1}X)$$

$$= X'V^{-1}E(HV^{-1}H)V^{-1}X.$$

But $E(HV^{-1}H) = E(\hat{V}-V)V^{-1}(\hat{V}-V) = E(\hat{V}V^{-1}\hat{V}) - V$.

Applying Lemma 2.2.3 with $U = V^{-1}$,

$$E(HV^{-1}H) = VV^{-1}V + 2\ \mathrm{diag}\left(\frac{\sigma_{ii}I_{n_i}}{n_i}\right) - V$$

$$E(HV^{-1}H) = 2\ \mathrm{diag}\left(\frac{\sigma_{ii}I_{n_i}}{n_i}\right) \tag{2.3.12}$$

$$\therefore E(A_2) = 2X'V^{-1}\left[\mathrm{diag}\left(\frac{\sigma_{ii}I_{n_i}}{n_i}\right)\right]V^{-1}X.$$

By (2.2.21) this becomes

$$E(A_2) = 2xL'V^{-1}\ \mathrm{diag}\left(\frac{\sigma_{ii}I_{n_i}}{n_i}\right)\ V^{-1}Lx'$$

$$\therefore E(A_2) = 2x\ \mathrm{diag}(1/\sigma_{ii})x' \tag{2.3.13}$$

$$E(A_1A_0^{-1}A_1) = E[X'V^{-1}HV^{-1}XA_0^{-1}X'V^{-1}HV^{-1}X].$$

By Lemma 2.2.9, this reduces to

$$E(A_1 A_0^{-1} A_1) = E(x h_\Delta x' A_0^{-1} x h_\Delta x') \tag{2.3.14}$$

Applying Lemma 2.2.3 with U as $x'A_0x$ = G(say),

we obtain

$$E(A_1 A_0^{-1} A_1) = 2x \operatorname{diag}\left(\frac{n_i g_i}{\sigma_{ii}^2}\right) x', \tag{2.3.15}$$

where $g_i = \underline{x}_i' A_o^{-1} \underline{x}_i$ as for Lemma 2.2.7.

Substituting (2.3.13) and (2.3.15) into (2.3.10) we get

$$\text{Term II} = A_0^{-1} + 2A_0^{-1}x \operatorname{diag}(1/\sigma_{ii}) x' A_0^{-1} - 2A_0^{-1} x \operatorname{diag}(n_i g_i/\sigma_{ii}^2) x' A_0^{-1}. \tag{2.3.16}$$

Now we shall evaluate some terms for Term I (2.3.11)

$$\begin{aligned} E(A_0^{-1}X'V^{-1}QV^{-1}XA_0^{-1}) &= A_0^{-1}X'E(V^{-1}QV^{-1})XA_0^{-1} \\ &= A_0^{-1}X'V^{-1}E(\underline{\epsilon\epsilon}'-V)V^{-1}XA_0^{-1} \\ &= A_0^{-1}X'V^{-1}(V-V)V^{-1}XA_0^{-1} = 0 \end{aligned} \tag{2.3.17}$$

$$E(A_0^{-1}X'V^{-1}HV^{-1}QV^{-1}XA_0^{-1}) = A_0^{-1}X'V^{-1}E(HV^{-1}Q)V^{-1}XA_0^{-1}$$

using Lemma 2.2.5

$$\begin{aligned} &= A_0^{-1}X'V^{-1}(2VDV^{-1}MV)V^{-1}XA_0^{-1} \\ &= 2A_o^{-1}X'DV^{-1}MXA_0^{-1} \end{aligned} \tag{2.3.18}$$

Similarly $E(A_0^{-1}X'V^{-1}QV^{-1}HV^{-1}XA_0^{-1}) = 2A_0^{-1}X'DV^{-1}MXA_0^{-1}$ (2.3.19)

$$E(A_0^{-1}A_1A_0^{-1}X'V^{-1}QV^{-1}XA_0^{-1}) = E(A_0^{-1}X'V^{-1}HV^{-1}XA_0^{-1}X'V^{-1}QV^{-1}XA_0^{-1})$$

Let $\quad T = V^{-1}XA_0^{-1}X'V^{-1}$

$$= E(A_0^{-1}X'V^{-1}HTQV^{-1}XA_0^{-1})$$

$$= A_0^{-1}X'V^{-1}E(HTQ)V^{-1}XA_0^{-1}$$

Applying Lemma 2.2.5

$$= A_0^{-1}X'V^{-1}(2VDT_\Delta V)V^{-1}XA_0^{-1}$$

$$= 2A_0^{-1}X'DT_\Delta XA_0^{-1} \tag{2.3.20}$$

Similarly

$$E(A_0^{-1}X'V^{-1}QV^{-1}XA_0^{-1}A_1A_0^{-1}) = 2A_0^{-1}X'DT_\Delta XA_0^{-1}\ . \tag{2.3.21}$$

Substituting the values of (2.3.17)-(2.3.20) in Equation (2.3.11) we obtain

$$\text{Term I} = 0-4A_0^{-1}X'DV^{-1}MXA_0^{-1}+4A_0^{-1}X'DT_\Delta XA_0^{-1}+0(n^{-3}) \tag{2.3.22}$$

where $\quad D = \begin{pmatrix} \frac{1}{n_1} I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{n_k} I_{n_k} \end{pmatrix}, \quad M = \begin{pmatrix} M_1 & & 0 \\ & \ddots & \\ 0 & & M_k \end{pmatrix}.$

and $\quad T_\Delta = \text{block diag}(V^{-1}XA_0^{-1}X'MV^{-1}). \qquad (2.3.23)$

Using Lemma 2.2.7, Equation (2.3.23) becomes

$$T_{\Delta} = V^{-1}\begin{pmatrix} g_1 M_{n_1} & & 0 \\ & \ddots & \\ 0 & & g_k M_{n_k} \end{pmatrix} V^{-1} \tag{2.3.24}$$

$$\therefore DT_{\Delta} = \begin{pmatrix} \frac{g_1 M_{n_1}}{n_1 \sigma_{11}^2} & & 0 \\ & \ddots & \\ 0 & & \frac{g_k M_{n_k}}{n_k \sigma_{kk}^2} \end{pmatrix} \tag{2.3.25}$$

Using Lemma 2.2.8

$$X'DV^{-1}MX = xL'DV^{-1}MLx'$$

$$= x\begin{pmatrix} n_1 \frac{1}{n_1\sigma_{11}} & & 0 \\ & \ddots & \\ 0 & & n_k \frac{1}{n_k\sigma_{kk}} \end{pmatrix} x' = 0 \tag{2.3.26}$$

Using the results of (2.3.25), (2.3.26), and (2.3.17) in (2.3.22)

$$\text{Term I} = 0 + 0(n^{-3}). \tag{2.3.27}$$

Substituting the results of (2.3.16) and (2.3.27) in (2.3.9), we get

$$\mathrm{Var}(\hat{\beta}_W) = A_0^{-1} + 2A_0^{-1}x \ \mathrm{diag}\{\sigma_{ii}^{-2}(\sigma_{ii} - n_i g_i)\}x'A_0^{-1} + 0(n^{-3}).$$

(ii) <u>The MINQUE Case</u>

Taking the linear model (2.2.1) $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$

The MINQU-based estimator for $\underline{\beta}$ is $\underline{\hat{\beta}}_M = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{Y}$

where
$$\hat{V} = \begin{bmatrix} \hat{\sigma}_{11} I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{kk} I_{n_k} \end{bmatrix}.$$

Using the MINQUE of $\hat{\sigma}_{ii}$ given in (2.2.8), it follows that, as before,

$$E(\hat{\underline{\beta}}_M) = \underline{\beta}_M.$$

Now we wish to evaluate

$$V(\hat{\underline{\beta}}_M) = E(\hat{\underline{\beta}}_M - \underline{\beta}_M)(\hat{\underline{\beta}}_M - \underline{\beta}_M)'$$

The results up to (2.3.11) are the same, so we will evaluate Terms I and II. Starting with Term II

$$E(A_2) = E[X'V^{-1}HV^{-1}HV^{-1}X] = X'V^{-1}E[HV^{-1}H]V^{-1}H$$

Applying Lemma 2.2.4 with $U = V^{-1}$, we get

$$E(HV^{-1}H) = \text{diag}\left[\frac{(C_{ii}+2W_{ii})}{\sigma_{ii}} I_{n_i}\right]$$

$$= 2\ \text{diag}\left(\frac{t_{ii}}{\sigma_{ii}} I_{n_i}\right) = 2T_\Delta \text{ say}$$

where t_{ii} is given in (2.2.15).

$$\therefore E(A_2) = 2X'V^{-1}T_\Delta V^{-1}X = 2x\ \text{diag}(n_i t_{ii}/\sigma_{ii}^3)x'. \qquad (2.2.28)$$

Likewise $E(A_1 A_0^{-1} A_1) = E[X'V^{-1}HV^{-1}XA_0^{-1}X'V^{-1}HV^{-1}X]$.

By Lemma 2.2.9, this becomes

$$E(A_1 A_0^{-1} A_1) = xE(h_\Delta x' A_0^{-1} x h_\Delta)x'.$$

Applying Lemma 2.2.4, it follows that

$$E(A_1A_0^{-1}A_1) = 2x(n_in_jt_{ij}g_{ij}/\sigma_{ii}^2\sigma_{jj}^2)x' \tag{2.2.29}$$

where $g_{ij} = \underline{x}_i'A_0^{-1}\underline{x}_j$ and t_{ij} is given by (2.2.16).

Substituting (2.3.28) and (2.3.29) into Term II (2.3.10)

$$\text{Term II} = A_0^{-1}+2A_0^{-1}x\ \text{diag}(n_it_{ii}/\sigma_{ii}^3)x'A_0^{-1}-2A_0^{-1}x(n_in_jt_{ij}g_{ij}/\sigma_{ii}^2\sigma_{jj}^2)x'A_0^{-1} \tag{2.3.30}$$

The results for Term I are as follows:

(i) $E(A_0^{-1}X'V^{-1}QV^{-1}XA_0^{-1}) = 0$; by (2.3.17) (2.3.31)

Using Lemma 2.2.6, the remaining terms of Term I ((2.3.18)-2.3.21)) will be as follows:

$$E(A_0^{-1}X'V^{-1}HV^{-1}QV^{-1}XA_0^{-1}) = 2A_0^{-1}X'CDV^{-1}MXA_0^{-1} \tag{2.3.32}$$

$$E(A_0^{-1}X'V^{-1}QV^{-1}HV^{-1}XA_0^{-1}) = 2A_0^{-1}X'CDV^{-1}MXA_0^{-1} \tag{2.3.33}$$

$$E(A_0^{-1}A_1A_0^{-1}X'V^{-1}QV^{-1}XA_0^{-1}) = 2A_0^{-1}X'CDT_\Delta XA_0^{-1} \tag{2.3.34}$$

and $E(A_0^{-1}X'V^{-1}QV^{-1}XA_0^{-1}A_1A_0^{-1}) = 2A_0^{-1}X'CDT_\Delta XA_0^{-1}$ (2.3.35)

where $T_\Delta = \text{block diag}(V^{-1}XA_0^{-1}X'MV^{-1})$. (2.3.36)

Substituting the values of (2.3.31)-(2.3.35) in (2.3.11),

$$\text{Term I} = 4A_0^{-1}X'CDT_\Delta XA_0^{-1}-4A_0^{-1}X'CDV^{-1}MXA_0^{-1}+0(n^{-3}) \tag{2.3.37}$$

where $D = \begin{pmatrix} \frac{1}{n_1} I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{n_k} I_{n_k} \end{pmatrix}$ and $M = \begin{pmatrix} M_1 & & 0 \\ & \ddots & \\ 0 & & M_k \end{pmatrix}$.

Using Lemma 2.2.7, (2.3.36) becomes

$$T_\Delta = V^{-1}\begin{pmatrix} g_1 M_{n_1} & & 0 \\ & \ddots & \\ 0 & & g_k M_{n_k} \end{pmatrix} V^{-1}$$

$$DT_\Delta = \begin{pmatrix} g_1 M_{n_1}/n_1\sigma_{11}^2 & & 0 \\ & \ddots & \\ 0 & & g_k M_{n_k}/n_k\sigma_{kk}^2 \end{pmatrix}$$

$$CDT_\Delta = \frac{n}{(n-2)}\begin{pmatrix} g_1 M_{n_1}/n_1^2\sigma_{11}^2 & & 0 \\ & \ddots & \\ 0 & & g_k M_{n_k}/n_k^2\sigma_{kk}^2 \end{pmatrix}$$

$$- \frac{1}{(n-1)(n-2)}\begin{pmatrix} (n_1-1)g_1 M_{n_1}/n_1\sigma_{11}^2 & & 0 \\ & \ddots & \\ 0 & & (n_k-1)g_k M_{n_k}/n_k^2\sigma_{kk}^2 \end{pmatrix}. \qquad (2.3.38)$$

Using Lemma 2.2.8 ($ML = 0$), the term

$$X'CDV^{-1}MX = xL'CDV^{-1}MLx' = 0. \qquad (2.3.39)$$

Thus, using the results of (2.3.38) and (2.3.39) in (2.3.37)

$$\text{Term I} = 0 + 0(n^{-3}).$$

So $\operatorname{Var}(\hat{\underline{\beta}}_M)$ = Term I + Term II, yields

$$\operatorname{Var}(\hat{\underline{\beta}}_M) = A_0^{-1} + 2A_0^{-1}x\ \operatorname{diag}(n_i t_{ii}/\sigma_{ii}^3)x'A_0^{-1} - 2A_0^{-1}x(n_i n_j t_{ij} g_{ij}/\sigma_{ii}^2\sigma_{jj}^2)x'A_0^{-1} + 0(n^{-3}).$$

2.4 Non-Normal Case

Theorem 2.4.1

Consider the linear model

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon}$$

where $\underline{Y}$ is a vector of $n = \sum_{i=1}^{k} n_i$ observations, X is a known matrix of order $n \times p$, $\underline{\beta}$ is a vector of p unknown parameters and $\underline{\epsilon}$ is a vector of n. Also $\epsilon_{ij} \sim ID(\underline{0}, \sigma_{ii}, \gamma_i)$ where γ_i is the measure of kurtosis. It is assumed that the distribution of the ϵ_{ij} is symmetric.

(i) The WLS Case

As before, the estimator is $\hat{\underline{\beta}}_W = (X'\hat{V}^{-1}X)'X'\hat{V}^{-1}\underline{Y}$

$$\text{where} \quad \hat{V} = \begin{bmatrix} \hat{\sigma}_{11} I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{kk} I_{n_k} \end{bmatrix} . \quad \text{and} \quad \hat{\sigma}_{ii} = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2/(n_i - 1).$$

Then $E(\hat{\underline{\beta}}_W) = \underline{\beta}_W$, and

$$\text{Var}(\hat{\underline{\beta}}_W) = A_0^{-1} + A_0^{-1} x \ \text{diag}\{\sigma_{ii}^{-2}(\sigma_{ii} - n_i g_i)(2+\gamma_i)\} x' A_0^{-1} + 0(n^{-3}). \quad (2.4.1)$$

(ii) MINQUE Case

As before, the estimator is $\hat{\underline{\beta}}_M = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{Y}$

$$\text{where} \quad \hat{V} = \begin{bmatrix} \hat{\sigma}_{11} I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{kk} I_{n_k} \end{bmatrix};$$

and the MINQUE of $\hat{\sigma}_{ii}$ is given in (2.2.8). Then $E(\hat{\underline{\beta}}_M) = \underline{\beta}_M$ and

$$\mathrm{Var}(\hat{\underline{\beta}}_M) = A_0^{-1}+2A_0^{-1}x\ \mathrm{diag}(n_i t_{ii}/\sigma_{ii}^3)x'A_0^{-1}-2A_0^{-1}x\left\{\frac{n_i n_j t_{ij} g_{ij}}{\sigma_{ii}^2\sigma_{jj}^2}\right\}x'A_0^{-1}+0(n^{-3}) \tag{2.4.2}$$

where, again by routine calculation,

$$t_{ii} = \mathrm{Var}(\hat{\sigma}_{ii}) = \frac{2}{(n-2)^2}\left[\sigma_{ii}^2(2+\gamma_i)\left\{\frac{n}{n_i}(n-4) + \frac{2(n+1)}{(n-1)}\right\}+n\sigma_{ii}\bar{\sigma}\left\{\frac{2+\gamma_i}{n_i} - \frac{4}{n-1}\right\}\right.$$

$$\left. + \frac{1}{2(n-1)^2}\left\{2n^2\bar{\sigma}^2-\Sigma n_i(2+\gamma_i)\sigma_{ii}^2\right\}\right] \tag{2.4.3}$$

and

$$t_{ij} = \mathrm{Cov}(\hat{\sigma}_{ii},\hat{\sigma}_{jj}) = 2(n-2)^{-2}\left[2\sigma_{ii}\sigma_{jj} + \frac{\sigma_{ii}^2(2+\gamma_i)+\sigma_{jj}^2(2+\gamma_i)}{(n-1)} - \frac{2n\bar{\sigma}}{n-1}(\sigma_{ii}+\sigma_{jj})\right.$$

$$\left. + \frac{1}{2}(n-1)^{-2}(2n^2\bar{\sigma}^2-\Sigma n_i(2+\gamma_i)\sigma_{ii}^2)\right]. \tag{2.4.4}$$

In order to prove Theorem 2.4.1, we first prove the following two lemmas:

<u>Lemma 2.4.1</u>

When $\hat{V}$ is estimated by WLS

$$E(\hat{V}U\hat{V}) = VUV + \mathrm{diag}\left(\frac{\sigma_{ii}^2 u_{ii}(2+\gamma_i)}{n_i}\right),$$

where U is any known matrix and

$$\hat{V} = \begin{bmatrix} \hat{\sigma}_{11} I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \sigma_k^2 I_{n_k} \end{bmatrix} .$$

Proof:

$$\hat{V}U\hat{V} \text{ has } (ij)^{th} \text{ element} = \sum_{kr}\sum \hat{\sigma}_{ik} u_{kr} \hat{\sigma}_{rj} \tag{2.4.5}$$

where $\hat{V} = \operatorname{diag}(\hat{\sigma}_{ii})$ $\quad \hat{\sigma}_{ii} = (n_i-1)^{-1}\Sigma(X_{ij}-\bar{X}_i)^2$

and $\hat{\sigma}_{ij} \equiv \sigma_{ij} \equiv 0; \quad i \neq j.$

$$E(\hat{\sigma}_{ii}) = \sigma_{ii}$$

Now,

$$\operatorname{Var}(\hat{\sigma}_{ii}) = \left(\frac{n_i}{n_i-1}\right)^2 \left[\frac{(n_i-1)^2}{n_i^3} \mu_{4i} - \frac{(n_i-1)(n_i-3)}{n_i^3} \mu_{2i}^2\right]$$

and $\quad E(\hat{\sigma}_{ii}^2) = \operatorname{Var}(\hat{\sigma}_{ii}) + \sigma_{ii}^2 = \dfrac{\mu_{4i}}{n_i} - \dfrac{(n_i-3)\sigma_{ii}^2}{(n_i-1)n_i} + \sigma_{ii}^2$,

where $\mu_{4i} = (\gamma_i+3)\sigma_{ii}^2$

$$E(\hat{\sigma}_{ii}^2) = \sigma_{ii}^2\left[1 + \frac{3+\gamma_i}{n_i} - \frac{(n_i-3)}{n_i(n_i-1)}\right]$$

$$= \sigma_{ii}^2\left[1 + \frac{\gamma_i}{n_i} + \frac{2}{(n_i-1)}\right]$$

$$\simeq \sigma_{ii}^2\left[1 + \frac{2+\gamma_i}{n_i}\right] .$$

Further, $E(\hat{\sigma}_{ii}\hat{\sigma}_{jj}) = \sigma_{ii}\sigma_{jj}$ and all other terms in (2.4.5) are equal to zero. Thus $E(\hat{V}U\hat{V})$ has the $(ij)^{th}$ element $= \sum_{k=i}\sum_{r=j} E(\hat{\sigma}_{ii}u_{ij}\hat{\sigma}_{jj})$

$$= \sigma_{ii}u_{ij}\sigma_{jj} \qquad \text{if } i\neq j$$

$$= \sigma_{ii}^2 u_{ii}\left(1 + \frac{2+\gamma_i}{n_i}\right) \qquad \text{where } i=j.$$

$$\therefore E(\hat{V}U\hat{V}) = VUV + \text{diag}\left(\frac{\sigma_{ii}^2 u_{ii}(2+\gamma_i)}{n_i}\right) . \qquad (2.4.6)$$

Lemma 2.4.2

When $\hat{V}$ is estimated by MINQU,

$$E(\hat{V}U\hat{V}) = VUV + B_1 + B_2$$

where $B_1 = \{u_{ij}C_{ij}\}$, $B_2 = \text{diag}\left\{\frac{u_{ii}W_{ii}(2+\gamma_i)}{n_i}\right\}$,

U is a known matrix and

$$\hat{V} = \begin{bmatrix} \hat{\sigma}_{11}I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{kk}I_{n_k} \end{bmatrix} ;$$

$\hat{\sigma}_{ii}$ being the MINQU estimator is given in (2.2.8).

Proof:

The proof is on the same lines as for Lemma 2.4.1.

$E(\hat{V}U\hat{V})$ has diagonal elements $E(U_{ii}\hat{\sigma}_{ii}^2)$ and off-diagonal elements $E(u_{ij}\hat{\sigma}_{ii}\hat{\sigma}_{jj})$. Further,

$$\hat{V} = \text{diag}(\hat{\sigma}_{ii}),\ E(\hat{\sigma}_{ii}) = \sigma_{ii} \quad \text{and}$$

$$\hat{\sigma}_{ij} \equiv \sigma_{ij} \equiv 0; \quad i \neq j.$$

It follows that

$$E(u_{ii}\hat{\sigma}_{ii}^2) = u_{ii}\sigma_{ii}^2 + u_{ii}\ \text{Var}(\hat{\sigma}_{ii}) \left(1 + \frac{2+\gamma_i}{n_i}\right)$$

and
$$E(u_{ij}\hat{\sigma}_{ii}\hat{\sigma}_{jj}) = u_{ij}\sigma_{ii}\sigma_{jj} + \text{Cov}(\hat{\sigma}_{ii}, \hat{\sigma}_{jj}).$$

Using the values of (2.4.3) and (2.4.4), we obtain

$$E(\hat{V}U\hat{V}) = VUV + B_1 + B_2. \tag{2.4.7}$$

(i) Proof of Theorem 2.4.1 (WLS Case)

Given the symmetry of the $\underline{\epsilon}_i$ distribution it follows by the same arguments that

$$E(\hat{\underline{\beta}}_W) = \hat{\underline{\beta}}_W.$$

For the variance, the results up to Equation (2.3.11) are the same. We now evaluate Terms I and II. Starting with Term II,

$$E(A_2) = E[X'V^{-1}HV^{-1}HV^{-1}X] = X'V^{-1}E[HV^{-1}H]V^{-1}X$$

using Lemma 2.4.1, with $U = V^{-1}$,

$$E(HV^{-1}H) = \text{diag}\left(\frac{\sigma_{ii}(2+\gamma_i)}{n_i}\, I_{n_i}\right)$$

$$E(A_2) = X'V^{-1}\ \text{diag}\left[\frac{\sigma_{ii}(2+\gamma_i)}{n_i}\, I_{n_i}\right] V^{-1}X.$$

By (2.2.21) this becomes

$$E(A_2) = xL'V^{-1}\ \text{diag}\left(\frac{\sigma_{ii}(2+\gamma_i)}{n_i}\, I_{n_i}\right) V^{-1}Lx'$$

$$= x\ \text{diag}\left(\frac{2+\gamma_i}{\sigma_{ii}}\right) x'. \tag{2.4.8}$$

Again using Lemma 2.4.1, with $U = V^{-1}$ and Lemma 2.2.9, we obtain the expression for $E(A_1A_0^{-1}A_1)$ as follows

$$E(A_1A_0^{-1}A_1) = x\ \text{diag}\left[\frac{n_i g_i}{\sigma_{ii}^2}(2+\gamma_i)\right] x'. \tag{2.4.9}$$

In this case, all the terms of Term I vanish. Substituting the values of Terms I and II in (2.3.9), we obtain

$$\text{Var}(\hat{\underline{\beta}}_W) = A_0^{-1} + A_0^{-1}x\ \text{diag}\left[\frac{(\sigma_{ii}-n_i g_i)(2+\gamma_i)}{\sigma_{ii}^2}\right] x'A_0^{-1} + 0(n^{-3}).$$

(ii) <u>Proof of Theorem 2.4.1 (MINQU Case)</u>

Given the symmetry of the $\underline{\epsilon}_i$ distribution it follows by the same arguments that

$$E(\hat{\underline{\beta}}_M) = \underline{\beta}_M\ .$$

For the variance, the results up to (2.3.11) are the same and the terms of Term I vanish by arguments similar to those in (2.3.31)-(2.3.39). So

$$\mathrm{Var}(\hat{\underline{\beta}}_M) = \text{Term II.} \tag{2.4.10}$$

Now we evaluate Term II.

$$E(A_2) = E[X'V^{-1}HV^{-1}HV^{-1}X] = X'V^{-1}E(HV^{-1}H)V^{-1}X.$$

Using Lemma 2.4.2 with $U = V^{-1}$

$$E(HV^{-1}H) = \mathrm{diag}\left[\frac{2t_{ii}}{\sigma_{ii}} I_{n_i}\right]$$

$$\therefore E(A_2) = 2x\ \mathrm{diag}[n_i t_{ii}/\sigma_{ii}^3]x' \tag{2.4.11}$$

where the value of t_{ii} is given in (2.4.3).

Likewise, applying Lemma 2.2.9 and Lemma 2.4.2,

$$E(A_1A_0^{-1}A_1) = 2x[n_i n_j t_{ij} g_{ij}/\sigma_{ii}^2\sigma_{jj}^2]x', \tag{2.4.12}$$

where t_{ij} is given in (2.4.4).

Substituting the values of (2.4.11) and (2.4.12) into (2.4.10), we obtain

$$\mathrm{Var}(\hat{\underline{\beta}}_M) = A_0^{-1}+2A_0^{-1}x\ \mathrm{diag}(n_i t_{ii}/\sigma_{ii}^3)x'A_0^{-1}-2A_0^{-1}x\left[\frac{n_i n_j t_{ij} g_{ij}}{\sigma_{ii}^2\sigma_{jj}^2}\right]x'A_0^{-1}+0(n^{-3}). \tag{2.4.13}$$

2.5 Special Cases of the Theorems

In order to examine the effect of the second order terms, we consider the following special case.

$$p = 1, \quad x = (1,1,\dots,1),$$

$$n_i = n/k \quad \text{and} \quad \sigma_{jj} = \sigma_{ii} = \sigma^2 \qquad \text{for } i=1,2,\dots,k.$$

It follows that

$$A_0 = X'V^{-1}X = xL'V^{-1}Lx' = x\begin{pmatrix} n_1/\sigma_{11} & & 0 \\ & \ddots & \\ 0 & & n_k/\sigma_{kk} \end{pmatrix} x' = \sum_{i=1}^{k} n_i/\sigma_{ii} = n/\sigma^2.$$

Hence $\qquad g_i = x'A_0^{-1}x = \sigma^2/n.$

Now we evaluate the variances for each case.

<u>Example 2.5.1</u>

Using WLS and assuming normality

$$V(\hat{\underline{\beta}}_W) = A_0^{-1} + 2A_0^{-1}x \,\mathrm{diag}\left\{\frac{1}{\sigma_{ii}^2}(\sigma_{ii} - n_i g_i)\right\} x'A_0^{-1} + 0(n^{-3})$$

$$= \frac{\sigma^2}{n} + 2\,\frac{\sigma^4 k}{n^2}\left\{\frac{1}{\sigma^2} - \frac{1}{k\sigma^2}\right\} + 0(n^{-3})$$

$$= \frac{\sigma^2}{n} + \frac{2k\sigma^2}{n^2}\left\{\frac{k-1}{k}\right\} + 0(n^{-3})$$

$$= \frac{\sigma^2}{n}\left\{1 + \frac{2(k-1)}{n}\right\} + 0(n^{-3}). \qquad (2.5.1)$$

Example 2.5.2

Using MINQUE and assuming normality

$$\mathrm{Var}(\hat{\beta}_M) = A_0^{-1}+2A_0^{-1}x\ \mathrm{diag}(n_i t_{ii}/\sigma_{ii}^3)x'A_0^{-1}-2A_0^{-1}x(n_i n_j t_{ij} g_{ij}/\sigma_{ii}^2\sigma_{jj}^2)x'A_0^{-1}+0(n^{-3})$$

where t_{ii} and t_{ij} are given in (2.2.15) and (2.2.16).

This becomes

$$\mathrm{Var}(\hat{\beta}_M) = \frac{\sigma^2}{n} + \frac{2k\sigma^4}{n^2}\left[\frac{n}{k}\left\{\frac{2k\sigma^4}{n-2} - \frac{2\sigma^4}{(n-1)(n-2)}\right\}/\sigma^6\right] - \frac{2\sigma^4}{n^2}\cdot\frac{\sigma^2}{n}\cdot\frac{n^2}{k^2}\cdot\frac{2\sigma^4}{\sigma^8}$$

$$\left[\frac{k\cdot k}{n-2} - \frac{k^2}{(n-1)(n-2)}\right] + 0(n^{-3})$$

$$= \frac{\sigma^2}{n} + \frac{4k\sigma^2}{n(n-2)} - \frac{4\sigma^2}{n(n-1)(n-2)} - \frac{4\sigma^2}{n(n-2)} + \frac{4\sigma^2}{n(n-1)(n-2)} + 0(n^{-3})$$

$$= \frac{\sigma^2}{n}\left[1 + \frac{4(k-1)}{(n-2)}\right] + 0(n^{-3}). \qquad (2.5.2)$$

Example 2.5.3

Using WLS without assuming normality and taking all $\gamma_i = \gamma$,

$$\mathrm{Var}(\hat{\beta}_W) = A_0^{-1}+A_0^{-1}x\ \mathrm{diag}\left[\sigma_{ii}^{-2}(\sigma_{ii}-n_i g_i)(2+\gamma_i)\right]x'A_0^{-1}+0(n^{-3})$$

$$= \frac{\sigma^2}{n} + (2+\gamma)\,\frac{\sigma^4 k}{n^2}\left[\frac{1}{\sigma^2} - \frac{1}{k\sigma^2}\right] + 0(n^{-3})$$

$$= \frac{\sigma^2}{n} + (2+\gamma)\,\frac{\sigma^2}{n^2}\,(k-1) + 0(n^{-3})$$

$$= \frac{\sigma^2}{n}\left(1 + \frac{(2+\gamma)(k-1)}{n}\right) + 0(n^{-3}). \qquad (2.5.3)$$

Example 2.5.4

Using MINQUE without assuming normality and taking all $\gamma_i = \gamma$,

$$Var(\hat{\beta}_M) = A_0^{-1}+2A_0^{-1}x\ \mathrm{diag}(n_i t_{ii}/\sigma_{ii}^3)x'A_0^{-1}-2A_0^{-1}x[n_i n_j t_{ij} g_{ij}/\sigma_{ii}^2\sigma_{jj}^2]x'A_0^{-1}+0(n^{-3})$$

where t_{ii} and t_{ij} are given in (2.4.3) and (2.4.4) respectively. It follows that

$$Var(\hat{\beta}_M) = \frac{\sigma^2}{n}\left[1 + \frac{(k-1)(4+\gamma)}{n-2}\right] + 0(n^{-3}). \qquad (2.5.4)$$

It is apparent that the MINQU-based estimators have higher variances, result which is further explored in Chapter 3.

CHAPTER 3

AN EMPIRICAL STUDY OF WLS AND MINQU-BASED ESTIMATORS FOR SMALL SAMPLES

3.1 Introduction

In this chapter, we will present a comparison of the theoretical results of Chapter 2 with simulated small sample results. These simulated and theoretical results are based on WLS and MINQU-based estimators.

We present results based on both normal and non-normal distributions (e.g., logistic and rectangular). Simulated results for WLS and MINQU-based estimators for the Cauchy are also presented. A Monte Carlo comparison of the modified MINQU-based estimators due to J. N. K. Rao (1973), OLS and WLS estimators is also given.

3.2 Sampling Experiments and Computations

For this investigation, the model of Jacquez, et al. (1968) is considered in which α and β are assumed to be equal to one. The model is

$$y_{ij} = 1 + x_i + \epsilon_{ij} \qquad i=1,2,\ldots,K;\ j = 1,2,\ldots,n_i$$

where the ϵ_{ij} are normally distributed with zero mean and variance σ_i^2. The values of K are chosen to be 4, 6, 8, and 10 and the corresponding chosen values for x_i are (1,4,7,10), (1,2,4,7,9,10), (1,2,4,5,6,7,9,10) and (1,2,3,4,5,6,7,8,9,10) respectively. The n_i's are chosen to be equal to m and the values of m considered are 2,3,4,5,10. For each (m,K) pair, three σ-patterns are chosen:

(1) $\sigma_i = 1$

(2) $\sigma_i = (x_i+8)/9$,

and (3) $\sigma_i = (0.5x_i+1)/3$.

We generated 200 samples for each set of (m,K) and σ-pattern for normal, logistic and rectangular and Cauchy distributed error terms. On the basis of these generated data, we compared the means and variances for $\hat{\underline{\beta}}$ based upon OLS, WLS, MINQU-based and modified MINQU-based estimators.

Computations were carried out on the Burroughs 6700 at Warwick University and the IBM 3033 at The Pennsylvania State University, USA

3.3 Empirical Results

The results are now presented in tabular form. In each table we give the variance of $\hat{\beta}$ based upon the 200 simulated samples and the approximate large sample variance. The ratio of the two is also given. Separate entries are given for m = 2,3,4,5 and 10 for both the WLS and MINQU-based estimators. The table numbers are as follows:

Values of σ_i	Distribution		
	Normal	Logistic	Rectangular
1	3.1	3.4	3.7
$(x_i+8)/9$	3.2	3.5	3.8
$(0.5x_i+1)/3$	3.3	3.6	3.9

The ratios of the simulated to asymptotic results are then summarized for both types of estimator in Tables 3.10-3.12. Biases of $\hat{\beta}$ for OLS, WLS, MINQU-based and J. N. K. Rao's Modified MINQU-based estimators for normal and non-normal (logistic and rectangular) distributions are presented in Tables 3.13-3.15. In Tables 3.16-3.18, we present a simulation-based comparison of the variances of the OLS, WLS, MINQU-based and Modified MINQU-based estimators. The results are evaluated in Section 3.4.

3.4 Results and Discussion

3.4.1 Choice of Estimators. The first and most striking conclusion to be drawn from the study is that the MINQU-based estimators are not suitable for this type of study. In all cases the variances of such estimators are much higher than those based on OLS and WLS. Indeed, the small sample results are even more striking than the asymptotic values. If anything, the MINQUE method is even worse than these tables suggest, since there were occasions when one or more of the variance estimators were negative so that the regression parameters could not be estimated.

Although the modified MINQUE is better than MINQUE, β still has an unacceptably large bias in both normal and non-normal cases and for all different patterns of σ_i, particularly for small values of m. All the bias values for $\hat{\beta}$ under WLS appear acceptable. The biases for $\hat{\alpha}$ in both normal and non-normal cases with different σ_i follow the same pattern.

The reason for the poor performance of the MINQU estimators would seem to be that the method "stretches" the variance estimates, or increases the differences between them. It would appear that a method of "shrinkage", as in Stein's method for several means, might perform better. We explore this possibility in Chapter 4 by a prior likelihood method.

3.4.2 Performance of Large Sample Approximation. As m increases the large samples results and the computer results generally draw closer together. However, the computer estimates of the variances are sometimes lower than we would expect, even if the σ_i were known. The reason for this discrepancy is not clear, but may reflect some slight deficiency in the random number generator. In general, the asymptotic results appear to be satisfactory for m=10. For smaller m they are less accurate but still provide a better estimate of the variance than the usual WLS expression assuming σ's known.

The approximation assumes that K is fixed whereas m becomes large. Therefore, it is not surprising that large discrepancies arise from m=2 and K=8 or 10. Rather the surprise is that the WLS variances are typically quite close for $m \geq 3$.

3.4.3 Effects of Non-normality. Variance ratios of $\hat{\beta}$ of asymptotic results show that in all cases WLS has better performance than MINQU-based and Modified MINQU-based estimators in normal and non-normal cases under all different patterns of errors. These asymptotic results also show that in non-normal cases, large γ gives larger variances of $\hat{\beta}$.

Although the theoretical variance for $\hat{\beta}$ does not exist for the Cauchy, the computer simulations produce a very heavy tailed distribution with finite variance. Thus the results in Table 3.19 confirm the supremacy of WLS over MINQUE for heavy tails, often by an extensive amount.

Table 3.1 Comparative variances of $\hat{\beta}$ for WLS and MINQU-based estimators when the errors are normally distributed with $\sigma_i = 1$, all i (all variances X100).

	WLS			MINQUE		
m	Computer Results	Theoretical Results	Ratio (%)	Computer Results	Theoretical Results	Ratio (%)
		K=4			K=4	
2	1.586	2.222	71.38			
3	1.059	1.234	85.82	3.259	3.704	87.99
4	0.585	0.833	70.23	2.189	1.666	131.39
5	0.527	0.622	84.73	1.355	1.037	130.67
10	0.288	0.267	107.86	0.407	0.334	121.86
∞		2.222/m			2.222/m	
		K=6			K=6	
2	2.119	1.439	147.26			
3	0.738	0.799	92.37	1.768	2.398	73.73
4	0.487	0.540	90.19	1.242	1.829	67.91
5	0.412	0.403	102.42	0.689	0.671	102.68
10	0.181	0.173	104.62	0.243	0.216	112.50
∞		1.439/m			1.439/m	
		K=8			K=8	
2	2.595	1.429	181.60			
3	0.916	0.794	115.37	93.912	2.381	3944.23
4	0.682	0.536	127.24	9.554	1.071	892.06
5	0.420	0.400	105.00	0.572	0.667	85.76
10	0.175	0.172	101.74	0.261	0.215	121.40
∞		1.429/m			1.429/m	
		K=10			K=10	
2	2.105	1.212	173.68			
3	0.840	0.673	124.81	3.020	2.020	149.50
4	0.507	0.455	111.43	1.398	0.909	153.79
5	0.325	0.339	95.87	0.616	0.566	108.83
10	0.147	0.145	101.38	0.180	0.182	98.90
∞		1.212/m			1.212/m	

Note: With m=2, MINQUE does not exist.

Table 3.2 Comparative variances of $\hat{\beta}$ for WLS and MINQU-based estimators when the errors are normally distributed with $\sigma_i = (x_i+8)/9$, all i (all variances X100).

	WLS			MINQUE		
m	Computer Results	Theoretical Results	Ratio (%)	Computer Results	Theoretical Results	Ratio (%)
		K=4			K=4	
2	3.707	4.726	78.44			
3	2.229	2.626	84.88	6.876	7.877	87.29
4	1.253	1.772	70.71	4.341	3.545	122.45
5	1.066	1.323	80.57	2.684	2.205	121.72
10	0.579	0.567	102.12	1.817	0.709	115.23
∞		4.726/m			4.726/m	
		K=6			K=6	
2	4.452	3.187	139.69			
3	1.590	1.771	89.78	3.481	5.312	65.53
4	1.127	1.195	94.31	2.587	2.390	108.24
5	0.887	0.892	99.44	1.461	1.487	98.25
10	0.411	0.382	107.59	0.602	0.478	125.94
∞		3.187/m			3.187/m	
		K=8			K=8	
2	5.466	2.993	182.63			
3	1.909	1.663	114.79	51.984	4.988	1042.18
4	1.465	1.122	130.57	2.648	2.245	117.95
5	1.071	0.838	127.80	1.763	1.397	126.20
10	0.373	0.359	103.90	0.550	0.449	122.49
∞		2.993/m			2.993/m	
		K=10			K=10	
2	4.660	2.558	182.17			
3	1.750	1.421	123.15	5.500	4.263	129.02
4	1.180	0.959	123.04	2.611	1.919	136.06
5	0.889	0.716	124.16	1.500	1.194	125.63
10	0.309	0.307	100.65	0.490	0.384	127.60
∞		2.558/m			2.558/m	

Note: With m=2, MINQUE does not exist.

Table 3.3 Comparative variances of β for WLS and MINQU-based estimators when the errors are normally distributed with $\sigma_i = (0.5x_i+1)/3$ (all variances X100).

	WLS			MINQUE		
m	Computer Results	Theoretical Results	Ratio (%)	Computer Results	Theoretical Results	Ratio (%)
		K=4			K=4	
2	2.650	2.903	91.28			
3	1.330	1.613	82.46	5.300	4.838	109.55
4	0.830	1.089	76.22	3.470	2.177	159.39
5	0.710	0.813	87.33	2.210	1.355	163.10
10	0.382	0.348	109.77	0.850	0.435	195.40
∞		2.903/m			2.903/m	
		K=6			K=6	
2	2.730	2.108	129.38			
3	1.120	1.171	95.73	2.830	3.513	80.56
4	0.810	0.791	102.53	2.180	1.581	137.89
5	0.632	0.590	107.12	1.190	0.984	120.93
10	0.269	0.253	106.32	0.560	0.316	177.22
∞		2.108/m			2.108/m	
		K=8			K=8	
2	3.130	1.703	184.12			
3	1.160	0.946	122.11	14.210	2.838	500.70
4	0.870	0.639	135.94	1.020	1.277	79.87
5	0.547	0.477	114.68	0.630	0.795	79.24
10	0.209	0.204	102.45	0.310	0.255	121.57
∞		1.703/m			1.703/m	
		K=10			K=10	
2	3.010	1.476	203.38			
3	1.110	0.820	135.37	3.300	2.460	134.15
4	0.720	0.554	129.97	1.930	1.107	174.35
5	0.526	0.413	127.36	0.820	0.689	119.01
10	0.186	0.177	105.08	0.301	0.221	136.20
∞		1.476/m			1.476/m	

Note: With m=2, MINQUE does not exist.

Table 3.4 Comparative variances of β for WLS and MINQU-based estimators when errors are logistically distributed with $\sigma_i = 1$, all i (all variances X100).

m	WLS Computer Results	WLS Theoretical Results	WLS Ratio (%)	MINQUE Computer Results	MINQUE Theoretical Results	MINQUE Ratio (%)
		K=4			K=4	
2	8.367	9.505	88.02			
3	4.950	5.037	98.27	10.959	15.107	72.54
4	3.284	3.290	99.82	5.262	6.580	79.97
5	2.265	2.398	94.45	2.718	3.997	68.01
10	1.201	0.964	124.59	1.449	1.207	120.05
∞		7.310/m			7.310/m	
		K=6			K=6	
2	5.360	6.155	87.08			
3	3.001	3.260	92.05	7.550	9.784	77.16
4	1.970	2.132	92.40	3.140	4.260	73.71
5	1.500	1.553	96.59	1.650	2.589	63.73
10	0.661	0.625	105.76	0.983	0.780	126.03
∞		4.734/m			4.734/m	
		K=8			K=8	
2	6.175	6.113	101.01			
3	2.789	3.237	86.16	203.085	9.715	2090.43
4	1.905	2.115	90.07	3.070	4.231	72.56
5	1.427	1.543	92.48	1.856	2.569	72.24
10	0.675	0.622	108.52	0.697	0.776	89.82
∞		4.701/m			4.701/m	
		K=10			K=10	
2	4.951	5.185	95.49			
3	2.553	2.747	92.94	10.608	8.241	128.72
4	1.735	1.793	96.76	3.541	3.589	98.66
5	1.350	1.309	103.13	1.746	2.181	80.05
10	0.539	0.526	102.47	0.607	0.658	92.25
∞		3.987/m			3.987/m	

Note: With m=2, MINQUE does not exist.

Table 3.5 Comparative variances of β for WLS and MINQU-based estimators when errors are logistically distributed with $\sigma_i = (x_i+8)/9$, all i (all variances X100).

	WLS			MINQUE		
m	Computer Results	Theoretical Results	Ratio (%)	Computer Results	Theoretical Results	Ratio (%)
		K=4			K=4	
2	19.381	20.213	95.88			
3	9.209	10.711	85.98	22.166	32.132	68.98
4	6.717	6.997	96.00	17.209	13.992	122.99
5	4.498	4.711	95.48	5.500	8.501	64.70
10	2.201	2.066	106.53	3.881	2.566	151.25
∞		15.548/m			15.548/m	
		K=6			K=6	
2	14.226	13.630	104.37			
3	6.179	7.221	85.57	16.808	21.667	77.57
4	4.690	4.718	99.41	7.209	9.435	76.41
5	3.520	3.438	102.39	4.866	5.731	84.91
10	1.413	1.385	102.02	0.917	1.730	53.01
∞		10.485/m			10.485/m	
		K=8			K=8	
2	11.877	12.801	92.78			
3	6.801	6.784	100.25	318.765	20.351	1566.34
4	3.770	4.431	85.10	16.780	8.860	189.39
5	3.165	3.231	97.96	4.916	5.382	91.34
10	1.352	1.299	104.08	1.389	1.625	85.48
∞		9.847/m			9.847/m	
		K=10			K=10	
2	9.988	10.939	91.31			
3	5.517	5.797	95.17	19.278	17.394	110.83
4	3.743	3.787	98.84	4.066	7.573	53.69
5	2.806	2.760	101.67	3.717	4.599	80.82
10	1.128	1.112	101.44	1.830	1.388	131.84
∞		8.416/m			8.416/m	

Note: With m=2, MINQUE does not exist.

Table 3.6 Comparative variances of β for WLS and MINQU-based estimators when errors are logistically distributed with $\sigma_i = (0.5x_i+1)/3$, all i (all variances X100).

	WLS			MINQUE		
m	Computer Results	Theoretical Results	Ratio (%)	Computer Results	Theoretical Results	Ratio (%)
		K=4			K=4	
2	10.303	12.416	82.98			
3	5.412	6.580	82.25	7.811	19.739	39.57
4	3.897	4.297	90.69	5.033	8.597	58.54
5	2.727	3.132	87.07	2.322	5.221	82.78
10	1.281	1.260	101.67	2.209	1.576	140.16
∞		9.551/m			9.551/m	
		K=6			K=6	
2	7.811	9.014	86.65			
3	4.799	4.777	100.46	4.325	14.334	72.03
4	2.908	3.122	93.14	2.970	6.248	79.54
5	2.275	2.273	100.09	2.707	3.790	71.42
10	0.943	0.915	103.06	0.698	1.145	60.96
∞		6.935/m			6.935/m	
		K=8			K=8	
2	7.138	7.284	98.00			
3	3.363	3.859	87.15	241.060	11.580	2081.69
4	2.240	2.520	88.89	2.890	5.043	57.31
5	1.999	1.839	108.70	1.455	3.063	47.50
10	0.761	0.740	102.84	1.690	0.925	182.70
∞		5.603/m			5.603/m	
		K=10			K=10	
2	5.887	6.311	93.28			
3	3.120	3.346	93.25	27.781	10.034	276.87
4	2.010	2.184	92.03	3.212	4.369	73.52
5	1.637	1.592	102.83	2.009	2.655	75.67
10	0.656	0.642	102.18	1.311	0.803	163.26
∞		4.855/m			4.855/m	

Note: With m=2, MINQUE does not exist.

Table 3.7 Comparative variances of β for WLS and MINQU-based estimators when errors are rectangularly distributed with $\sigma_i=1$ (all variances X100).

	WLS			MINQUE		
m	Computer Results	Theoretical Results	Ratio (%)	Computer Results	Theoretical Results	Ratio (%)
		K=4			K=4	
2	0.149	0.130	114.62			
3	0.089	0.078	114.10	0.309	0.235	131.49
4	0.058	0.056	103.57	0.088	0.111	79.30
5	0.045	0.043	104.65	0.059	0.072	81.94
10	0.020	0.020	100.00	0.021	0.025	84.00
∞		0.185/m			0.185/m	
		K=6			K=6	
2	0.100	0.084	119.05			
3	0.058	0.051	113.73	0.247	0.152	162.50
4	0.042	0.036	116.67	0.053	0.072	73.61
5	0.027	0.028	96.42	0.037	0.046	80.43
10	0.014	0.013	107.69	0.015	0.016	93.75
∞		0.120/m			0.120/m	
		K=8			K=8	
2	0.099	0.083	119.28			
3	0.062	0.050	124.00	10.510	0.151	6960.26
4	0.043	0.036	119.44	0.076	0.071	107.04
5	0.030	0.028	107.14	0.037	0.046	80.43
10	0.013	0.013	100.00	0.014	0.016	87.50
∞		0.119/m			0.119/m	
		K=10			K=10	
2	0.098	0.071	138.03			
3	0.058	0.043	134.88	0.075	0.128	58.59
4	0.035	0.030	116.67	0.073	0.061	119.67
5	0.026	0.023	113.04	0.033	0.039	84.62
10	0.012	0.011	109.09	0.015	0.014	107.14
∞		0.101/m			0.101/m	

Note: With m=2, MINQUE does not exist.

Table 3.8 Comparative variances of β for WLS and MINQU-based estimators when errors are rectangularly distributed with $\sigma_i = (x_i+8)/9$, all i (all variances X100).

	WLS			MINQUE		
m	Computer Results	Theoretical Results	Ratio (%)	Computer Results	Theoretical Results	Ratio (%)
		K=4			K=4	
2	0.361	0.276	130.80			
3	0.179	0.166	107.83	0.251	0.499	50.30
4	0.104	0.118	88.14	0.173	0.236	73.31
5	0.107	0.091	117.58	0.131	0.152	86.18
10	0.043	0.043	100.00	0.048	0.053	90.56
∞		0.394/m			0.394/m	
		K=6			K=6	
2	0.255	0.186	137.10			
3	0.138	0.112	123.21	0.208	0.336	61.90
4	0.096	0.080	120.00	0.102	0.159	64.15
5	0.063	0.062	101.61	0.080	0.103	77.67
10	0.029	0.029	100.00	0.033	0.036	91.66
∞		0.266/m			0.266/m	
		K=8			K=8	
2	0.201	0.175	114.86			
3	0.126	0.105	120.00	12.386	0.316	3919.62
4	0.087	0.075	116.00	0.116	0.150	77.33
5	0.060	0.058	103.45	0.086	0.096	89.58
10	0.026	0.027	96.30	0.029	0.034	85.29
∞		0.249/m			0.249/m	
		K=10			K=10	
2	0.148	0.149	99.33			
3	0.097	0.090	107.78	0.196	0.270	72.59
4	0.068	0.064	106.25	0.093	0.128	72.66
5	0.051	0.049	104.08	0.070	0.082	85.36
10	0.024	0.023	104.35	0.027	0.029	93.10
∞		0.213/m			0.213/m	

Note: With m=2, MINQUE does not exist.

Table 3.9 Comparative variances of β for WLS and MINQU-based estimators when errors are rectangularly distributed with $\sigma_i = (0.5x_i+1)/3$, all i (all variances X100).

	WLS			MINQUE		
m	Computer Results	Theoretical Results	Ratio (%)	Computer Results	Theoretical Results	Ratio (%)
		K=4			K=4	
2	0.223	0.169	131.95			
3	0.117	0.102	114.71	0.501	0.306	163.73
4	0.065	0.073	89.04	0.092	0.145	63.45
5	0.061	0.056	108.93	0.074	0.094	78.72
10	0.025	0.026	96.15	0.027	0.033	81.82
∞		0.242/m			0.242/m	
		K=6			K=6	
2	0.147	0.123	119.51			
3	0.080	0.074	108.11	0.347	0.223	155.61
4	0.059	0.053	111.32	0.081	0.105	77.14
5	0.042	0.041	102.44	0.053	0.068	77.94
10	0.019	0.019	100.00	0.020	0.024	83.33
∞		0.176/m			0.176/m	
		K=8			K=8	
2	0.110	0.099	111.11			
3	0.064	0.060	106.67	8.331	0.180	4628.33
4	0.051	0.043	118.60	0.099	0.085	116.47
5	0.034	0.033	103.03	0.041	0.055	74.55
10	0.014	0.015	93.33	0.014	0.019	73.68
∞		0.142/m			0.142/m	
		K=10			K=10	
2	0.091	0.086	105.81			
3	0.065	0.052	125.00	0.131	0.156	83.97
4	0.041	0.037	110.81	0.054	0.074	72.97
5	0.030	0.029	103.45	0.052	0.048	108.33
10	0.014	0.013	107.69	0.015	0.017	88.24
∞		0.123/m			0.123/m	

Note: With m=2, MINQUE does not exist.

Table 3.10 Comparison of ratios of simulated to asymptotic variances for WLS and MINQU-based estimators of β for normal, logistic and rectangular errors for $\sigma_i = 1$, all i (results are in percent).

	WLS			MINQUE		
m	Normal	Logistic	Rectangular	Normal	Logistic	Rectangular
		K=4			K=4	
2	71.38	88.02	114.62			
3	85.82	98.27	114.10	87.99	72.54	131.49
4	70.23	99.82	103.57	131.39	79.97	79.30
5	84.73	94.45	104.65	130.67	68.01	81.94
10	107.86	124.59	100.00	121.86	120.05	84.00
		K=6			K=6	
2	147.26	87.08	119.05			
3	92.37	92.05	113.73	73.73	77.16	162.50
4	90.19	92.40	116.67	67.91	73.71	73.61
5	102.42	96.59	96.42	102.68	63.73	80.43
10	104.62	105.76	107.69	112.50	126.03	93.75
		K=8			K=8	
2	181.60	101.01	119.28			
3	115.37	86.16	124.00	3944.23	2090.43	6960.26
4	127.24	90.07	119.44	892.06	72.56	107.04
5	105.00	92.48	107.14	85.76	72.24	80.43
10	101.74	108.52	100.00	121.40	89.82	87.50
		K=10			K=10	
2	173.68	95.49	138.03			
3	124.81	92.94	134.88	149.50	128.72	58.59
4	111.43	96.76	116.67	153.79	90.66	119.67
5	95.87	103.13	113.04	108.83	80.05	84.62
10	101.38	102.47	109.09	98.90	92.25	107.14

Note: With m=2, MINQUE does not exist.

Table 3.11 Comparison of ratios of simulated to asymptotic variances for WLS and MINQU-based estimators of β for normal, logistic, and rectangular errors for $\sigma_i = (x_i+8)/9$, all i (results are in percent).

m	WLS Normal	WLS Logistic	WLS Rectangular	MINQUE Normal	MINQUE Logistic	MINQUE Rectangular
		K=4			K=4	
2	78.44	95.88	130.80			
3	84.88	85.98	107.83	87.29	68.98	50.30
4	70.71	96.00	88.14	122.45	122.99	73.31
5	80.57	95.48	117.58	121.72	64.70	86.18
10	102.12	106.53	100.00	115.23	151.25	90.56
		K=6			K=6	
2	139.69	104.37	137.10			
3	89.78	85.57	123.21	65.53	77.57	61.90
4	94.31	99.41	120.00	108.24	76.41	64.15
5	99.44	102.39	101.61	98.25	84.91	77.67
10	107.59	102.02	100.00	125.94	53.01	91.66
		K=8			K=8	
2	182.63	92.78	114.86			
3	114.79	100.25	120.00	1042.18	1566.34	3919.62
4	130.57	85.10	116.00	117.95	189.39	77.33
5	127.80	97.96	103.45	126.20	91.34	89.58
10	103.90	104.08	96.30	122.49	85.48	85.29
		K=10			K=10	
2	182.17	91.31	99.33			
3	123.15	95.17	107.78	129.02	110.83	72.59
4	123.04	98.84	106.25	136.06	53.69	72.66
5	124.16	101.67	104.08	125.63	80.82	85.36
10	100.65	101.44	104.35	127.60	131.84	93.10

Note: With m=2, MINQUE does not exist.

Table 3.12 Comparison of ratios of simulated to asymptotic variances for WLS and MINQU-based estimators of β for normal, logistic, and rectangular errors for $\sigma_i = (0.5x_i+1)/3$, all i (results are in percent).

	WLS			MINQUE		
m	Normal	Logistic	Rectangular	Normal	Logistic	Rectangular
		K=4			K=4	
2	91.28	82.98	131.95			
3	82.46	82.25	114.71	109.55	39.57	163.73
4	76.22	90.69	89.04	159.39	58.54	63.45
5	87.33	87.07	108.93	163.10	82.78	78.72
10	109.77	101.67	96.15	195.40	140.16	81.82
		K=6			K=6	
2	129.38	86.65	119.51			
3	95.73	100.46	108.11	80.56	72.03	155.61
4	102.53	93.14	111.32	137.89	79.54	77.14
5	107.12	100.09	102.44	120.93	71.42	77.94
10	106.32	103.06	100.00	177.22	60.96	83.33
		K=8			K=8	
2	184.12	98.00	111.11			
3	122.11	87.15	106.67	500.70	2081.69	4628.33
4	135.94	88.89	118.60	79.87	57.31	116.47
5	114.68	108.70	103.03	79.24	47.50	74.55
10	102.45	102.84	93.33	121.57	182.70	73.68
		K=10			K=10	
2	203.38	93.28	105.81			
3	135.37	93.25	125.00	134.15	276.87	83.97
4	129.97	92.03	110.81	174.35	73.52	72.97
5	127.36	102.83	103.45	119.01	75.67	108.33
10	105.08	102.18	107.69	136.20	163.26	88.24

Note: With m=2, MINQUE does not exist.

Table 3.13 Empirical biases of β for OLS, WLS, MINQU-based and Modified MINQU-based estimators when errors are normally distributed (biases are in percent).

		$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
m	K →	4	6	8	10	4	6	8	10	4	6	8	10
2	OLS	0.901	0.390	0.104	-0.105	1.495	-0.334	0.116	0.200	7.700	8.870	11.010	11.800
	WLS	-0.215	-0.791	-2.148	-0.594	-0.998	-1.206	-2.627	-0.800	-1.100	-0.850	-1.670	-0.600
	MINQ	-11.677	-16.837	0.037	-30.900	-19.993	-28.846	-47.418	-18.760	-17.230	-23.640	150.450	-3.450
	MMIN	-10.788	-10.888	-9.220	-18.240	-8.680	-19.010	-23.190	-7.866	-12.240	-9.380	31.210	-2.789
3	OLS	-0.281	-0.078	1.099	0.423	-0.566	-0.309	1.215	0.288	2.140	6.730	9.800	10.180
	WLS	-0.348	-0.496	1.151	-0.092	0.160	-0.314	1.877	-0.491	0.510	0.080	1.550	-0.500
	MINQ	-7.015	-6.955	-10.414	-18.528	-11.808	-12.428	-0.754	-19.590	-9.650	-10.270	-17.950	-19.470
	MMIN	-10.620	-5.879	-3.679	-6.700	-12.348	-10.610	-1.789	-12.390	0.730	-8.730	31.000	-17.724
4	OLS	-0.525	0.353	0.495	0.367	-1.263	0.326	0.870	0.396	-1.370	4.820	8.390	9.920
	WLS	-0.924	0.619	0.813	0.394	-1.060	0.602	0.869	0.300	0	0.270	0.400	0.640
	MINQ	-5.639	-4.683	-10.226	-7.559	-9.504	-8.566	-14.308	-11.630	-8.170	-7.500	-14.370	-12.610
	MMIN	-4.829	0.003	-8.679	8.211	-5.448	-6.230	-7.710	-2.433	-5.730	34.000	9.320	-4.340
5	OLS	1.374	-0.028	0.007	0.017	1.910	0.149	0.131	-0.100	1.490	2.370	6.570	8.130
	WLS	1.407	-0.002	-0.017	0.326	1.672	0.014	0.095	0.301	1.000	0.070	0.200	8.130
	MINQ	-2.153	-3.953	-7.156	-7.406	-4.191	-6.634	-8.599	-10.310	-3.580	-5.410	-7.800	-9.570
	MMIN	-1.239	0.895	-2.239	-6.390	-6.509	-2.000	-7.088	-11.771	-2.190	-6.120	-8.870	-3.450
10	OLS	0.221	-0.153	0.002	0.185	0.469	-0.267	-0.043	0.066	0.500	-0.250	0	0.010
	WLS	0.160	-0.282	0.041	0.306	0.462	-0.258	0.100	0.371	0.500	-0.090	0.050	0.170
	MINQ	-0.600	-1.657	-3.700	-2.579	-1.120	-3.060	-4.580	-4.090	-0.700	-0.250	-4.290	-3.520
	MMIN	-2.800	2.351	32.110	-3.640	-2.680	-2.980	-3.000	-3.780	-1.870	-1.210	-1.790	9.890

Table 3.14 Empirical biases of β for OLS, WLS, MINQU-based and Modified MINQU-based estimators when errors are logistically distributed (biases are in percent).

		$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
m	K →	4	6	8	10	4	6	8	10	4	6	8	10
2	OLS	1.384	-0.390	0.156	2.278	1.407	-2.677	0.911	3.121	-1.777	-0.241	1.271	3.012
	WLS	3.994	-0.387	-1.281	1.206	-2.801	1.022	-1.099	1.606	-2.678	1.881	0.822	1.009
	MINQ	-18.694	-28.418	-9.657	-29.683	-19.620	-31.622	-40.117	-18.187	-10.653	-90.620	-67.206	-8.613
	MMIN	-7.976	-27.332	-10.871	-17.789	-8.607	-25.186	-9.633	-29.123	-6.912	-32.001	-29.127	-3.393
3	OLS	2.035	-0.257	-0.925	-0.840	2.557	3.110	-1.000	1.001	1.993	0.506	-1.613	1.231
	WLS	2.137	-0.687	0.082	-1.775	2.609	-1.821	0.082	-1.666	2.107	-1.072	-1.029	-0.987
	MINQ	-11.806	-14.887	-9.893	-21.386	-7.691	-6.711	-7.222	-18.621	-9.662	-32.621	-7.875	-17.680
	MMIN	-10.756	-9.212	-6.796	-27.100	-3.210	-6.222	-5.137	-13.370	-8.772	-20.210	-6.000	-11.777
4	OLS	-1.593	1.137	-0.517	-0.741	-0.389	0.633	0.687	-0.667	-1.286	0.007	-0.999	1.000
	WLS	-1.836	0.633	-0.999	-0.637	-0.699	-0.002	-1.000	-0.554	-1.711	0.086	-1.001	1.006
	MINQ	-10.245	-7.906	-13.723	-15.571	-5.245	-3.591	-10.109	-9.960	-14.071	-21.003	-9.136	-6.960
	MMIN	-3.713	-2.810	-9.888	-10.132	-2.121	-3.612	-7.812	-6.712	-10.090	-12.811	-4.286	-3.999
5	OLS	-0.441	-0.814	0.510	-0.094	0.226	-1.007	-0.350	0.229	2.321	1.393	1.003	-1.967
	WLS	-0.335	-0.002	0.581	-0.471	0.009	0.012	-0.981	0.828	1.667	0.621	1.321	0.890
	MINQ	-7.109	-8.471	-8.816	-9.913	-2.661	-6.603	-9.621	-4.297	-9.686	-10.007	-3.210	-6.006
	MMIN	-6.622	-4.797	-5.697	-8.812	-2.991	3.133	-7.661	-3.212	-8.786	-9.116	-2.001	-5.177
10	OLS	-0.396	0.714	-0.101	0.701	-0.923	0.937	-1.283	-0.807	-1.015	0.920	-1.021	-1.065
	WLS	-0.610	0.489	-0.255	0.452	-1.211	1.110	-1.780	-0.411	-0.994	0.069	-0.621	-0.009
	MINQ	-3.882	-3.591	-5.068	-4.627	-6.895	-3.210	-3.521	-2.603	-6.036	-4.081	-6.070	-3.292
	MMIN	3.110	1.031	-4.889	-3.609	-7.121	2.880	-2.912	-3.101	-4.137	2.099	-4.112	-2.690

Table 3.15 Empirical biases of β for OLS, WLS, and MINQU-based estimators when errors are rectangularly distributed (biases are in percent).

		$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
m	K →	4	6	8	10	4	6	8	10	4	6	8	10
	OLS	1.879	-1.097	0.017	0.023	3.375	-1.084	-1.760	-1.211	1.785	-1.900	0.977	0.239
2	WLS	-2.680	-1.266	6.694	3.110	0.012	-2.701	9.969	-0.560	-2.529	-1.979	7.053	1.200
	MINQ	-25.668	-19.479	-43.493	-17.166	4.801	-290.865	-52.108	-71.411	-13.667	-26.117	-47.760	-37.810
	OLS	2.364	0.660	2.377	1.552	3.322	-0.241	3.181	-4.528	1.999	0.595	2.228	-2.227
3	WLS	4.462	1.679	4.105	3.102	5.993	1.883	4.182	-0.773	3.877	1.976	3.501	-0.811
	MINQ	-24.100	-39.356	-52.245	-41.661	-19.832	-34.561	-71.412	-45.693	-12.901	-28.250	-76.712	-37.127
	OLS	1.285	-0.454	0.264	0.099	1.417	0.407	0.382	4.321	1.877	-0.798	0.987	1.176
4	WLS	1.079	-0.615	-0.077	1.172	0.370	-1.089	1.127	4.515	1.899	-0.990	-0.007	1.188
	MINQ	-17.303	-23.490	-23.362	-19.910	-15.196	-25.096	-27.478	-25.076	-9.789	-17.251	-14.745	-19.199
	OLS	-0.567	0.311	2.239	0.001	-1.644	0.006	3.806	2.445	0.009	1.003	1.986	1.999
5	WLS	-0.168	-0.193	1.922	0.807	2.715	-0.825	2.257	2.590	-1.007	-0.007	2.200	2.003
	MINQ	-13.487	-14.444	-16.295	-17.710	-11.071	-16.002	-16.274	-21.303	-8.870	-9.687	-15.111	-16.109
	OLS	-0.543	0.916	-0.015	-0.009	-1.015	1.439	0.324	1.269	-1.006	0.882	-0.039	-0.002
10	WLS	-0.100	0.302	-0.263	-0.100	-0.994	0.320	0.043	1.900	-0.999	0.600	-0.172	-0.091
	MINQ	-5.535	-7.528	-8.114	-6.660	-6.036	-8.316	-8.582	-10.831	-4.076	-8.107	-7.210	-8.032

Table 3.16 Estimated variances for the OLS, WLS, MINQU-based and Modified MINQU-based estimators under the normal distribution (all variances X100).

		$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
m	K →	4	6	8	10	4	6	8	10	4	6	8	10
2	OLS	1.221	0.743	0.626	0.602	2.917	1.491	1.806	1.410	2.46	1.120	1.050	0.987
	WLS	1.586	2.119	2.595	2.105	3.707	4.452	5.466	4.660	2.650	2.730	3.130	3.010
3	OLS	0.778	0.492	0.484	0.404	1.971	1.427	1.682	1.307	1.440	0.940	0.976	0.860
	WLS	1.059	0.738	0.916	0.840	2.229	1.590	1.909	1.750	1.330	1.120	1.160	1.010
	MINQ	3.259	1.768	93.912	9.020	6.876	3.481	251.984	5.500	5.300	2.830	314.210	3.300
	MMIN	2.909	1.780	17.711	7.916	3.321	3.511	139.211	7.711	4.770	1.906	57.120	1.229
4	OLS	0.558	0.375	0.386	0.312	1.137	0.879	1.443	0.880	1.000	0.850	0.866	0.680
	WLS	0.585	0.487	0.682	0.507	1.253	1.227	1.465	1.180	0.830	0.810	0.870	0.720
	MINQ	2.189	1.242	39.554	3.398	4.341	2.587	2.648	2.611	3.470	2.180	1.020	1.930
	MMIN	2.311	1.250	3.578	3.786	3.110	2.232	2.780	3.119	2.866	2.005	1.112	2.500
5	OLS	0.451	0.289	0.294	0.269	1.030	0.945	0.940	0.896	0.880	0.670	0.580	0.563
	WLS	0.527	0.412	0.420	0.325	1.066	0.887	1.071	0.889	0.710	0.632	0.547	0.526
	MINQ	1.355	0.689	0.572	0.616	2.684	1.461	1.763	1.500	2.210	1.190	0.630	0.820
	MMIN	1.400	0.615	0.779	0.399	3.466	1.077	1.990	1.411	2.278	1.311	2.120	1.660
10	OLS	0.249	0.166	0.155	0.132	0.590	0.486	0.442	0.334	0.393	0.340	0.290	0.260
	WLS	0.288	0.181	0.175	0.147	0.579	0.411	0.373	0.309	0.382	0.270	0.209	0.186
	MINQ	0.407	0.243	0.261	0.180	0.817	0.602	0.550	0.490	0.850	0.560	0.310	0.301
	MMIN	0.377	0.303	0.207	0.226	1.001	0.669	0.639	0.490	1.001	0.422	0.219	0.267

Note: MINQUE with m=2 does not exist.

Table 3.17 Estimated variances for the OLS, WLS, MINQU-based and Modified MINQU-based estimatros under the logistic distribution (all variances X100).

		$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
m	K →	4	6	8	10	4	6	8	10	4	6	8	10
2	OLS	8.411	2.600	2.330	2.009	6.110	5.953	4.709	3.808	7.276	4.272	3.787	2.989
	WLS	0.367	5.360	6.175	4.951	9.381	14.226	11.877	9.988	10.303	7.811	7.138	5.887
3	OLS	3.300	2.791	2.011	1.333	4.291	5.301	4.746	3.879	5.357	4.798	3.353	2.710
	WLS	4.950	3.001	2.789	2.553	9.209	6.179	6.801	5.517	5.412	4.799	3.363	3.120
	MINQ	10.959	7.550	203.085	10.608	22.209	16.808	318.765	19.278	7.811	10.325	241.060	27.781
	MMIN	6.220	3.991	119.266	7.778	11.900	9.798	221.865	8.767	11.761	8.767	112.707	19.211
4	OLS	2.521	1.210	1.310	1.001	3.766	4.138	3.901	3.229	3.091	2.903	2.499	2.133
	WLS	3.284	1.970	1.905	1.735	6.717	4.690	3.770	3.743	3.897	2.908	2.240	2.010
	MINQ	5.262	3.140	3.070	3.541	17.209	7.209	16.780	4.066	5.033	2.970	2.890	3.212
	MMIN	5.101	2.077	4.020	2.321	10.100	4.989	11.210	3.880	4.806	2.667	2.910	2.888
5	OLS	1.803	0.933	0.917	1.008	2.888	2.760	2.872	2.888	2.920	2.343	1.301	1.666
	WLS	2.265	1.500	1.427	1.350	4.498	3.520	3.165	2.806	2.727	2.275	1.999	1.637
	MINQ	2.718	1.650	1.856	1.746	5.500	3.866	4.916	3.717	4.322	2.707	1.455	2.009
	MMIN	2.690	1.660	1.729	1.745	4.494	3.609	4.901	3.090	4.120	2.880	1.510	2.100
10	OLS	0.717	0.605	0.669	0.403	2.377	1.380	1.366	1.127	0.982	0.980	0.806	0.673
	WLS	1.201	0.661	0.675	0.539	2.201	1.413	1.352	1.128	1.281	0.943	0.761	0.656
	MINQ	1.449	0.983	0.697	0.607	3.881	0.917	1.389	1.830	2.209	0.698	1.690	1.311
	MMIN	1.510	1.001	0.580	0.501	2.110	1.101	1.370	1.711	2.031	0.711	0.979	0.960

Note: MINQUE with m=2 does not exist.

Table 3.18 Estimated variances for the OLS, WLS and MINQU-based estimators under the rectangular distribution (all variances X100).

		$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
m	K →	4	6	8	10	4	6	8	10	4	6	8	10
2	OLS	0.095	0.065	0.059	0.053	0.214	0.159	0.152	0.140	0.089	0.081	0.070	0.069
	WLS	0.149	0.100	0.099	0.098	0.361	0.255	0.201	0.148	0.147	0.110	2.775	0.091
3	OLS	0.056	0.032	0.040	0.038	0.138	0.077	0.101	0.072	0.077	0.066	0.063	0.056
	WLS	0.089	0.058	0.062	0.058	0.179	0.138	0.126	0.097	0.117	0.080	0.064	0.065
	MINQ	0.309	0.247	10.510	0.075	0.251	0.208	12.386	0.196	0.501	0.347	8.331	0.131
4	OLS	0.044	0.028	0.029	0.033	0.107	0.066	0.072	0.052	0.066	0.052	0.028	0.048
	WLS	0.058	0.042	0.043	0.035	0.104	0.096	0.087	0.068	0.065	0.059	0.051	0.041
	MINQ	0.088	0.053	0.076	0.073	0.173	0.102	0.116	0.093	0.092	0.081	0.099	0.054
5	OLS	0.035	0.018	0.027	0.025	0.088	0.046	0.065	0.050	0.047	0.029	0.027	0.029
	WLS	0.045	0.027	0.030	0.026	0.107	0.063	0.060	0.051	0.061	0.042	0.034	0.030
	MINQ	0.059	0.037	0.037	0.033	0.131	0.080	0.086	0.070	0.074	0.053	0.041	0.052
10	OLS	0.018	0.010	0.010	0.011	0.047	0.029	0.026	0.033	0.023	0.019	0.016	0.016
	WLS	0.020	0.014	0.013	0.012	0.043	0.029	0.026	0.024	0.025	0.017	0.014	0.014
	MINQ	0.021	0.015	0.014	0.015	0.048	0.033	0.029	0.027	0.027	0.020	0.014	0.015

Note: MINQUE with m=2 does not exist.

Table 3.19 Comparison of the simulated results of WLS and MINQU-based estimators for variances of β when errors are Cauchy distributed (all variances X100).

	$\sigma_i = 1$		$\sigma_i = (x_i+8)/9$		$\sigma_i = (0.5x_i+1)/9$	
m	WLS	MINQUE	WLS	MINQUE	WLS	MINQUE
	K=4		K=4		K=4	
2	8.580		14.256		9.751	
3	9.618	39.957	13.821	26.293	10.655	17.595
4	9.329	87.447	10.662	18.759	8.295	9.675
5	9.012	25.532	10.526	27.900	7.258	9.009
10	7.300	18.301	8.132	19.756	7.073	8.005
	K=6		K=6		K=6	
2	13.401		19.660		11.525	
3	2.863	18.291	6.910	9.757	4.282	11.753
4	1.963	8.776	4.206	6.898	3.121	9.295
5	3.072	6.770	4.896	5.991	2.899	7.085
10	2.791	6.246	3.217	4.875	2.695	4.818
	K=8		K=8		K=8	
2	9.207		15.696		10.851	
3	2.825	29.884	4.880	39.251	3.986	27.533
4	2.442	7.728	4.775	8.751	3.765	7.135
5	2.366	5.669	3.989	7.803	3.096	6.651
10	1.940	4.097	1.827	4.212	1.905	3.233
	K=10		K=10		K=10	
2	6.664		8.566		7.755	
3	2.059	6.244	3.765	7.756	3.600	6.597
4	1.525	4.064	3.002	7.212	2.898	4.788
5	1.793	3.368	2.465	5.751	2.907	4.252
10	1.414	2.454	2.009	3.009	1.882	2.288

CHAPTER 4

USE OF POSTERIOR LIKELIHOOD IN ESTIMATION

4.1 Introduction

In the previous chapter, a comparative empirical study of OLS, WLS, MINQU-based, and Modified MINQU-based estimators was presented. It was shown that MINQU-based or Modified MINQU-based estimators were not suitable for this type of study. The biases and variances of such estimators are much higher than those based on OLS and WLS. Also, estimators based on MINQUE may give rise to one or more negative variance estimates. One reason for the poor performance of the MINQU-based estimators would seem to be that the method "stretches" the variance estimators; that is, the differences between the estimates are increased relative to the usual sample variances.

In this chapter we will present a posterior likelihood (PL) method of estimation which "shrinks" the variance estimates rather than stretching them. The general method is described in Section 4.1.1. In Section 4.2, we derive the forms of the estimators under this approach and some theoretical results for $\underline{\hat{\beta}}_i$ and $\underline{\hat{\theta}}_i$ $(\hat{\theta}_i = \hat{\sigma}_i^2)$ are obtained, using Gamma prior likelihoods. In Section 4.3, the empirical comparative results of variances and biases of $\hat{\beta}$ for OLS, WLS, ML, and PL estimators will be presented. The conclusions are presented in Section 4.4.

<u>4.1.1 Prior Likelihood Estimation</u>. Consider the standard problem of estimating the unknown parameters $\underline{\theta}$ given sample data, $\underline{Y}$. In this

section we investigate the estimation of $\underline{\theta}$ under the assumption that prior likelihoods may be specified for the $\underline{\theta}$. That is, instead of maximizing the likelihood for the sample

$$\ell_S(\underline{\theta}|\underline{Y}) \tag{4.1.1}$$

we specify the prior likelihood

$$\ell_{prior}(\underline{\theta}|\underline{a}) \tag{4.1.2}$$

where $\underline{a}$ denotes a summary of initial information. We then maximize the posterior likelihood

$$\ell_{posterior}(\underline{\theta}|\underline{a},\underline{Y}) = \ell_{prior}(\underline{\theta}|\underline{a})\ell_S(\underline{\theta}|\underline{Y}). \tag{4.1.3}$$

At first sight this approach would seem to be equivalent to examination of the posterior mode in the Bayesian approach. However, there are several important differences.

First of all, the frequentist distinction between parameters and random variables is preserved; the final statement remains one concerning estimates of the parameters in light of the available data. This, in turn, implies that all the asymptotic properties of maximum likelihood estimators will continue to hold, provided that the prior likelihood is strictly positive over some open set of values which include the true parameter values.

Secondly, if there is no prior information available, then it is both sensible and feasible to set

$$\ell_{prior} = 1, \text{ for all } \underline{\theta} \in \Omega , \tag{4.1.4}$$

where Ω denotes the whole parameter space. Unlike the Bayesian approach, this specification of prior ignorance poses no difficulties since no integration over the parameter space is required.

Finally, we note that a transformation of the parameters will leave unchanged the posterior likelihood estimators, since there is no Jacobian involved.

Since the method of prior likelihood proceeds by specifying only an ordering of the likelihood function on the parameter space, further restrictions could be imposed to make this behave in a "Bayesian fashion". Thus, if we suppose that only a finite number of $\underline{\theta}$ values is of interest, $(\underline{\theta}_1,\ldots,\underline{\theta}_k)$ say then we could construct functions which satisfy the axioms of probability, namely

$$P^*_{\text{prior}}(\underline{\theta}_j|\underline{a}) = \frac{\ell_{\text{prior}}(\underline{\theta}_j|\underline{a})}{\sum_{i=1}^{k} \ell_{\text{prior}}(\underline{\theta}_i|\underline{a})} \tag{4.1.5}$$

and $$P^*_{\text{posterior}}(\underline{\theta}_j|\underline{a},\underline{Y}) = \frac{\ell_{\text{posterior}}(\underline{\theta}_j|\underline{a})}{\sum_{i=1}^{k} \ell_{\text{posterior}}(\underline{\theta}_i|\underline{a})} \ . \tag{4.1.6}$$

Further, it is readily shown that one may update from prior to posterior by use of Bayes' Theorem. These "probability functions" could be used, for example, in applications of decision theory and would produce equivalent results to the usual Bayesian approach.

In the next section we use the prior likelihood approach to develop estimators for the regression model.

4.2 Posterior Likelihood Estimators for the Regression Model

Consider the regression model

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon}$$

where $\underline{\epsilon}' = (\underline{\epsilon}_1',\ldots,\underline{\epsilon}_k')$ and $\underline{\epsilon}_i \sim IN(\underline{0},\theta_i I)$, $i=1,\ldots,k$. The standard assumptions are made concerning X and $\underline{\beta}$. In this section we investigate the estimation of $\underline{\beta}$ and $\underline{\theta} = (\theta_1,\ldots,\theta_k)'$ assuming that prior information on $\underline{\theta}$ may be expressed in the form of an incomplete gamma function; that is, for each θ_i

$$\theta_i^{-1} \sim \text{Independent gamma } (\alpha_i,\lambda_i),\ \alpha_i > 0,\ \lambda_i > 0.$$

No prior information is assumed for $\underline{\beta}$, so that

$$\ell_{prior}(\underline{\beta}) = 1.$$

The accuracy of these estimates will depend upon the accuracy of the prior likelihoods. Asymptotically, the resulting estimators $\hat{\beta}_p$ and $\hat{\theta}_p$ are the same as the WLS estimators when the parameters of the prior likelihoods (i.e., α_i and λ_i) are fixed.

4.2.1 Linear Regression Model with Different Prior Gamma Parameters.

Consider the regression model

$$\underline{Y}_i = X_i\underline{\beta} + \underline{\epsilon}_i \qquad (4.2.1)$$

Let $\underline{Y}_i \sim IN(X_i\underline{\beta},\theta_i I)$ $i=1,2,\ldots,k$ where $\underline{Y}_i$ is a vector of n_i observation, X_i is a known matrix of order $n_i \times p$, $\underline{\beta}$ is a vector of p unknown parameters and $\theta_i = \sigma_i^2$ denotes the i^{th} variance.

Let $\theta_i^{-1} \sim \text{Gamma}(\alpha_i,\lambda_i);\ \alpha_i > 0,\ \lambda_i > 0\ .$ (4.2.2)

Taking $\theta_i^{-1} \sim$ Gamma (α_i,λ_i) as a prior function for the above model (4.2.1)

$$\ell_{prior}(\underline{\theta}|\alpha_i,\lambda_i) = \text{Const} \prod_{i=1}^{k} \left[\theta_i^{-\lambda_i/2} e^{-\alpha_i/2\theta_i}\right] . \qquad (4.2.3)$$

The likelihood function for the sample is

$$\ell_s(\underline{\theta},\underline{\beta}|\underline{Y}_i,\alpha_i,\lambda_i) = \prod_{i=1}^{k} \theta_i^{-n_i/2} \exp\left\{-\frac{1}{2\theta_i}(\underline{Y}_i-X_i\underline{\beta})'(\underline{Y}_i-X_i\underline{\beta})\right\}. \qquad (4.2.4)$$

We obtain the posterior likelihood function by multiplying (4.2.3) and (4.2.4)

$$\ell_{posterior}(\underline{\theta},\underline{\beta}|\underline{Y}_i,\alpha_i,\lambda_i) = \prod_{i=1}^{K} \theta_i^{-\frac{1}{2}(n_i+\lambda_i)} \exp\left[-\frac{1}{2\theta_i}\{(\underline{Y}_i-X_i\underline{\beta})'(\underline{Y}_i-X_i\underline{\beta})+\alpha\}\right]$$

(4.2.5)

Taking log of (4.2.5) and differentiating w. r. to $\underline{\theta}$ and $\underline{\beta}$, the posterior likelihood estimators are

$$\hat{\theta}_i = \frac{\{(\underline{Y}_i-X_i\hat{\underline{\beta}})'(\underline{Y}_i-X_i\hat{\underline{\beta}})+\alpha_i\}}{n_i+\lambda_i} \quad i=1,2,\ldots,k \qquad (4.2.6)$$

$$\text{and} \quad \hat{\underline{\beta}} = \left[\sum_{i=1}^{k} X_i'X_i|\hat{\theta}_i\right]^{-1}\left[\sum_{i=1}^{k} X_i'\underline{Y}_i|\hat{\theta}_i\right] \qquad (4.2.7)$$

<u>4.2.2 Properties of PL Estimators.</u> First of all, we note that as $n_i \to \infty$ with α_i and λ_i fixed

$$\hat{\theta}_i \text{ approaches } \frac{(\underline{Y}_i - X_i\hat{\underline{\beta}})'(\underline{Y}_i - X_i\hat{\underline{\beta}})}{n_i} \tag{4.2.8}$$

and $\hat{\underline{\beta}}$ is given by (4.2.7), so that the estimators approach the regular ML solutions as we should expect. Thus, their large sample properties are equivalent to those of ML estimators.

Using (4.2.1)

$$\hat{\underline{\beta}} = (\Sigma X_i'X_i|\theta_i)^{-1}(\Sigma X_i'(X_i\underline{\beta}+\underline{\epsilon}_i)|\theta_i)$$

$$= \underline{\beta} + (\Sigma X_i'X_i|\theta_i)^{-1}\Sigma(X_i'\underline{\epsilon}_i|\theta_i) \tag{4.2.9}$$

Provided that the distribution of the ϵ_i is symmetric about zero it follows, by arguments similar to those of Section 2.2 (Lemma 2.2.1) that

$$E(\hat{\underline{\beta}}) = \underline{\beta} \quad . \tag{4.2.10}$$

That is, $\hat{\underline{\beta}}$ is unbiased.

The large sample variance for $\hat{\underline{\beta}}$ will be the same as for regular weighted least squares and for maximum likelihood, namely,

$$\text{Var}(\hat{\underline{\beta}}) = \sum_{i=1}^{K} (X_i'X_i|\theta_i)^{-1} \quad . \tag{4.2.11}$$

In the empirical work of Section 4.3, these estimators were simplified somewhat in that "posterior weighted least squares" was used. That is, for each i, if

$$n_i s_i^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2, \tag{4.2.12}$$

the posterior estimator for θ_i was taken as

$$\hat{\theta}_i = \frac{n_i s_i^2 + \alpha_i}{n_i + \lambda_i} \quad , \quad i=1,2,\ldots,k \qquad (4.2.13)$$

and $\hat{\underline{\beta}}$ was then obtained from (4.2.7).

In referring to the prior likelihood assumptions in the next section, the abbreviation (1,1) refers to an assumption of separate gamma densities for each θ_i with $\alpha=1$ and $\lambda=1$.

4.3 Empirical Results and Procedures

In order to study the properties of the posterior likelihood estimators emprically, the following procedure was adopted. The simple regression model of Jacquez, et al. (1968) was considered (as in Section 3.2), in which α and β are assumed to be equal to one. The values of k are chosen to be 4, 6, 8, and 10 and the corresponding chosen values for x_i are (1,4,7,10), (1,2,4,7,9,10), (1,2,4,5,6,7,9,10), and (1,2,3,4,5,6,7,8,9,10) respectively. The n_i are chosen to be equal to m and the values of m considered are 2, 3, 4, 5, 10. For each (m,k) pair, three sigma patterns are chosen: (1) $\sigma_i = 1$, (2) $\sigma_i = (x_i+8)/9$, and (3) $\sigma_i = (0.5x_i+1)/3$.

Two hundred simulated samples were drawn for each set of (m,k) and each σ-pattern. The results are summarized in tabular form. In the first three tables, the empirical variances of $\hat{\beta}$ for OLS, WLS, ML, and PL with different prior likelihoods {(1,1), (1,2), (2,2)} are presented for comparison; in these tables it was assumed that the ϵ_{ij} followed independent normal distributions. On the same lines Tables 4.4-4.9 are presented for non-normal (logistic and Cauchy distributions) cases.

The bias of $\hat{\beta}$ for OLS, WLS, ML, and PL with different prior likelihoods for normal and non-normal distributions are presented in Tables 4.10-4.12.

4.4 Results and Discussion

From the investigation described in Section 4.3 on the biases and variances of $\hat{\beta}$ the following conclusions can be drawn:

1) In the normal case, when the error variances σ_i^2 are equal, the OLS estimators provide more efficient results than WLS and ML for small m, as we would expect. As m increases (e.g., m=10), the results for the OLS and PL estimators become very close (see Table 4.1).

2) In the normal cases, when error variances are not equal, our simulated results (Tables 4.2 and 4.3) show that OLS are good for small samples and for small m (e.g., m=3) but as m increases and sample size 'k' increases, OLS estimators gradually get worse. The WLS and PL estimators are close up to m=5, but for m=10, PL shows its superiority over both WLS and ML.

3) In the non-normal cases, with equal error variances (see Tables 4.4 and 4.7), OLS again gives the best results. As m increases, the difference between OLS and PL gradually decreases and with m=10 and k=8 or 10 PL and OLS are very close.

4) With unequal error variances and non-normal errors, the performance of WLS and ML estimators are poor for small m. As m increases, the WLS and PL estimators start to provide the more efficient results. When m=10 and k=8 or 10, the PL estimators again show its superiority over WLS estimators.

5) In the Cauchy case, the prior "information" stabilizes the variance estimators and the PL method behaves considerably better than

WLS or ML for small m values, even when the information is incorrect.

6) Turning to the results on bias in Tables 4.10-4.12, all of these appear acceptably small although the ML method seems to generate somewhat larger biases for small m.

The most striking thing about the results for the PL estimators is that, even with different prior likelihoods, the method gives fairly similar and generally quite efficient results. These estimators show that the "shrinkage" of the variance differences appears to be beneficial, at least for samples with $m \leq 10$ replicates.

The minimum variance estimator for each combination of m, k, and σ is shown in Table 4.13; where two or more estimators are very close, both are listed.

Table 4.1 Estimated variances of $\hat{\beta}$ for the OLS, WLS, ML, and PL (with different prior assumptions) estimators when the errors are normally distributed with $\sigma_i = 1$, all i (all variances X100).

m	K=→	4	6	8	10
2	O	1.221	0.743	0.626	0.602
	W	1.586	2.119	2.595	2.105
	M	1.725	2.176	2.621	1.790
	P_1	1.226	0.857	1.020	0.745
	P_2	1.457	0.944	0.859	0.782
	P_3	1.050	0.936	0.717	0.581
3	O	0.778	0.492	0.484	0.404
	W	1.059	0.738	0.916	0.840
	M	1.148	0.821	0.924	0.992
	P_1	0.927	0.639	0.589	0.457
	P_2	0.756	0.619	0.507	0.488
	P_3	0.935	0.429	0.507	0.463
4	O	0.558	0.375	0.386	0.312
	W	0.585	0.487	0.682	0.507
	M	0.642	0.592	0.780	0.561
	P_1	0.606	0.393	0.434	0.312
	P_2	0.590	0.409	0.445	0.379
	P_3	0.620	0.399	0.417	0.343
5	O	0.451	0.289	0.294	0.269
	W	0.527	0.412	0.420	0.325
	M	0.546	0.399	0.345	0.398
	P_1	0.459	0.318	0.397	0.290
	P_2	0.457	0.316	0.303	0.347
	P_3	0.444	0.323	0.347	0.298
10	O	0.249	0.166	0.155	0.132
	W	0.288	0.181	0.175	0.147
	M	0.262	0.187	0.165	0.147
	P_1	0.223	0.179	0.165	0.137
	P_2	0.236	0.162	0.161	0.145
	P_3	0.202	0.170	0.145	0.102

O = OLS
W = WLS
M = ML
P_1 = PL(1,1)
P_2 = PL(1,2)
P_3 = PL(2,2)

Table 4.2 Estimated variances of $\hat{\beta}$ for the OLS, WLS, ML, and PL (with different prior assumptions) estimators when the errors are normally distributed with $\sigma_i = (x_i+8)/9$, all i (all variances X100).

m	K⇒	4	6	8	10
2	O	2.917	1.491	1.806	1.410
	W	3.707	4.452	5.466	4.660
	M	4.500	4.557	5.292	4.924
	P_1	3.604	2.442	2.434	2.058
	P_2	2.960	2.744	1.853	1.823
	P_3	3.543	1.873	2.083	1.621
3	O	1.971	1.427	1.682	1.307
	W	2.229	1.590	1.909	1.750
	M	2.891	1.963	1.762	2.110
	P_1	2.162	1.492	1.514	1.079
	P_2	2.157	1.125	1.585	1.258
	P_3	1.659	1.470	1.343	1.418
4	O	1.137	0.879	1.443	0.880
	W	1.253	1.127	1.465	1.180
	M	1.364	1.355	1.492	1.534
	P_1	1.578	1.042	1.018	0.835
	P_2	1.558	1.077	1.095	0.953
	P_3	1.414	1.035	0.960	0.960
5	O	1.030	0.945	0.940	0.896
	W	1.066	0.887	1.071	0.889
	M	1.214	0.837	1.060	0.812
	P_1	1.161	0.784	0.844	0.742
	P_2	1.202	0.752	0.766	0.645
	P_3	0.983	0.987	0.731	0.640
10	O	0.590	0.486	0.442	0.334
	W	0.579	0.411	0.373	0.309
	M	0.484	0.439	0.359	0.288
	P_1	0.495	0.368	0.315	0.223
	P_2	0.510	0.335	0.288	0.281
	P_3	0.533	0.367	0.228	0.258

O = OLS
W = WLS
M = ML
P_1 = PL(1,1)
P_2 = PL(1,2)
P_3 = PL(2,2)

Table 4.3 Estimated variances of $\hat{\beta}$ for the OLS, WLS, ML, and PL (with different prior assumptions) estimators when the errors are normally distributed with $\sigma_i = (0.5x_i+1)/3$, all i (all variances X100).

m	K→	4	6	8	10
2	O	2.460	1.120	1.050	0.987
	W	2.650	2.730	3.130	3.010
	M	2.911	2.810	3.201	3.206
	P_1	2.509	1.212	1.071	1.448
	P_2	2.313	1.367	0.986	1.511
	P_3	2.077	1.104	1.088	1.389
3	O	1.440	0.940	0.976	0.860
	W	1.330	1.120	1.160	1.110
	M	1.558	1.390	1.202	1.190
	P_1	1.274	1.064	0.989	0.872
	P_2	1.166	0.998	0.976	0.860
	P_3	1.192	0.967	0.968	0.870
4	O	1.000	0.850	0.866	0.680
	W	0.830	0.810	0.870	0.720
	M	0.912	0.906	0.923	0.898
	P_1	0.896	0.783	0.661	0.644
	P_2	0.845	0.792	0.695	0.696
	P_3	0.836	0.770	0.639	0.604
5	O	0.880	0.670	0.580	0.563
	W	0.710	0.632	0.547	0.526
	M	0.786	0.660	0.537	0.475
	P_1	0.760	0.691	0.510	0.396
	P_2	0.778	0.688	0.471	0.410
	P_3	0.742	0.693	0.480	0.382
10	O	0.393	0.340	0.290	0.260
	W	0.382	0.270	0.209	0.186
	M	0.383	0.276	0.212	0.160
	P_1	0.357	0.190	0.182	0.156
	P_2	0.360	0.181	0.178	0.150
	P_3	0.341	0.187	0.169	0.128

O = OLS
W = WLS
M = ML
P_1 = PL(1,1)
P_2 = PL(1,2)
P_3 = PL(2,2)

Table 4.4 Estimated variances of $\hat{\beta}$ for the OLS, WLS, ML, and PL (with different prior assumptions) estimators when the errors are logistically distributed with $\sigma_i = 1$, all i (all variances X100).

m	K=→	4	6	8	10
2	O	3.411	2.600	2.330	2.009
	W	8.367	5.360	6.175	4.951
	M	6.310	6.122	8.511	5.709
	P_1	5.350	3.316	3.906	3.510
	P_2	5.101	4.008	3.521	2.180
	P_3	4.189	2.806	2.800	2.623
3	O	3.300	2.791	2.011	1.333
	W	4.950	3.001	2.789	2.553
	M	3.301	2.713	2.591	2.910
	P_1	2.910	1.800	2.091	1.768
	P_2	3.501	2.082	2.007	1.680
	P_3	2.611	1.610	1.890	1.491
4	O	2.521	1.210	1.310	1.001
	W	3.284	1.970	1.905	1.735
	M	2.911	1.722	1.637	1.503
	P_1	2.500	1.286	1.365	1.181
	P_2	2.189	1.462	1.301	1.166
	P_3	1.932	1.263	1.468	1.260
5	O	1.803	0.933	0.917	1.008
	W	2.265	1.500	1.427	1.350
	M	1.880	1.417	1.228	1.009
	P_1	1.501	1.119	1.067	0.890
	P_2	1.807	1.101	0.810	0.901
	P_3	1.711	1.007	1.066	0.889
10	O	0.717	0.605	0.669	0.403
	W	1.201	0.661	0.675	0.539
	M	0.917	0.637	0.509	0.488
	P_1	0.767	0.359	0.363	0.481
	P_2	0.771	0.355	0.360	0.479
	P_3	0.689	0.361	0.361	0.482

O = OLS
W = WLS
M = ML
P_1 = PL(1,1)
P_2 = PL(1,2)
P_3 = PL(2,2)

Table 4.5 Estimated variances of $\hat{\beta}$ for the OLS, WLS, ML, and PL (with different prior assumptions) estimators when the errors are logistically distributed with $\sigma_i = (x_i+8)/9$, all i (all variances X100).

m	K=→	4	6	8	10
2	O	6.110	5.953	4.709	3.808
	W	19.381	14.226	11.877	9.988
	M	13.691	13.890	14.317	11.065
	P_1	9.082	8.066	6.781	6.665
	P_2	9.321	7.855	6.095	5.888
	P_3	9.003	6.797	5.139	6.014
3	O	4.291	5.301	4.746	3.879
	W	9.209	6.179	6.801	5.517
	M	8.311	7.808	6.913	6.713
	P_1	6.360	4.991	4.801	2.989
	P_2	6.711	4.336	4.550	3.121
	P_3	5.820	4.299	4.073	3.006
4	O	3.766	4.138	3.901	3.229
	W	6.717	4.690	3.770	3.743
	M	5.121	4.320	4.121	3.896
	P_1	4.800	3.221	3.011	2.447
	P_2	4.343	3.200	2.988	2.310
	P_3	4.009	3.168	3.107	2.401
5	O	2.888	2.660	2.872	2.888
	W	4.498	3.520	3.165	2.806
	M	4.717	3.990	3.107	2.411
	P_1	3.171	2.861	2.700	2.011
	P_2	3.404	2.636	2.580	1.896
	P_3	3.196	2.135	2.016	1.866
10	O	2.377	1.380	1.366	1.127
	W	2.201	1.413	1.352	1.128
	M	2.886	2.189	1.600	1.231
	P_1	1.766	1.110	0.912	0.907
	P_2	1.700	1.041	0.943	0.918
	P_3	1.722	1.018	0.939	0.925

O = OLS
W = WLS
M = ML
P_1 = PL(1,1)
P_2 = PL(1,2)
P_3 = PL(2,2)

Table 4.6 Estimated variances of $\hat{\beta}$ for the OLS, WLS, ML, and PL (with different prior assumptions) estimators when the errors are logistically distributed with $\sigma_i = (0.5x_i+1)/3$, all i (all variances X100).

m	K→	4	6	8	10
2	O	7.276	4.272	3.787	2.989
	W	10.303	7.811	7.138	5.887
	M	10.298	9.660	9.698	8.789
	P_1	7.889	5.197	4.391	4.200
	P_2	7.308	5.382	4.004	3.877
	P_3	5.910	4.990	4.009	3.816
3	O	5.357	4.798	3.353	2.710
	W	5.412	4.799	3.363	3.120
	M	5.960	5.800	5.163	4.969
	P_1	4.313	3.676	3.217	3.001
	P_2	4.213	3.389	3.100	3.077
	P_3	4.399	3.296	2.968	2.960
4	O	3.091	2.903	2.499	2.133
	W	3.897	2.908	2.240	2.010
	M	3.137	3.210	2.986	2.870
	P_1	3.066	2.510	2.100	2.009
	P_2	2.765	2.189	1.891	1.910
	P_3	2.688	2.006	1.888	1.766
5	O	2.920	2.343	1.301	1.666
	W	2.727	2.275	1.399	1.637
	M	2.391	2.088	1.965	1.882
	P_1	1.660	1.591	1.210	1.101
	P_2	1.697	1.500	1.207	1.007
	P_3	1.609	1.413	1.003	0.998
10	O	0.982	0.980	0.806	0.673
	W	1.281	0.943	0.761	0.656
	M	2.180	1.566	1.402	1.069
	P_1	0.977	0.711	0.498	0.480
	P_2	0.913	0.720	0.517	0.466
	P_3	0.944	0.638	0.510	0.479

O = OLS
W = WLS
M = ML
P_1 = PL(1,1)
P_2 = PL(1,2)
P_3 = PL(2,2)

Table 4.7 Estimated variances of $\hat{\beta}$ for the OLS, WLS, ML, and PL (with different prior assumptions) estimators when the errors are Cauchy distributed with σ_i = 1, all i (all variances X100).

m	K→	4	6	8	10
	O	7.861	4.213	3.003	2.336
	W	8.580	13.401	9.207	6.664
	M	13.132	14.053	8.942	6.961
2	P_1	21.685	4.601	3.237	2.219
	P_2	6.682	3.449	3.145	2.649
	P_3	11.947	4.612	3.456	2.741
	O	4.976	3.177	2.700	2.004
	W	9.618	2.863	2.825	2.059
	M	12.021	4.332	4.364	3.449
3	P_1	9.433	4.267	2.980	1.988
	P_2	8.476	3.753	2.822	1.905
	P_3	8.613	4.186	2.892	1.905
	O	4.009	2.001	2.239	1.432
	W	9.329	1.963	2.442	1.525
	M	7.680	2.819	1.931	1.885
4	P_1	12.259	2.749	2.227	1.560
	P_2	6.461	3.584	3.559	1.986
	P_3	7.827	2.531	2.147	2.071
	O	3.396	2.960	2.163	1.606
	W	9.012	3.072	2.366	1.793
	M	12.876	4.072	2.744	1.570
5	P_1	7.479	3.923	2.683	1.692
	P_2	7.601	2.273	2.439	1.413
	P_3	9.317	2.778	2.310	1.896
	O	3.189	2.661	1.951	1.438
	W	7.300	2.791	1.940	1.413
	M	6.724	3.032	1.923	1.340
10	P_1	5.830	3.350	1.794	1.306
	P_2	8.050	2.784	2.768	1.438
	P_3	7.626	2.459	2.280	1.350

O = OLS
W = WLS
m = MO
P_1 = PL(1,1)
P_2 = PL(1,2)
P_3 = PL(2,2)

Table 4.8 Estimated variances of $\hat{\beta}$ for the OLS, WLS, ML, and PL (with different prior assumptions) estimators when the errors are Cauchy distributed with $\sigma_i = (x_i+8)/9$, all i (all variances X100)

m	K→	4	6	8	10
	O	8.129	6.700	6.623	6.036
	W	11.977	14.866	12.163	9.375
	M	14.606	14.990	13.259	11.781
2	P_1	12.992	8.667	5.678	4.961
	P_2	9.398	6.282	6.037	4.877
	P_3	12.321	8.303	5.916	4.790
	O	9.070	4.088	3.881	3.520
	W	11.875	4.017	3.918	3.629
	M	13.710	6.221	5.477	4.711
3	P_1	11.891	5.733	4.301	3.212
	P_2	9.327	4.909	4.017	3.323
	P_3	9.550	5.207	4.003	3.470
	O	8.365	5.513	4.321	3.100
	W	10.712	3.399	3.360	2.418
	M	11.677	4.407	4.391	3.303
4	P_1	11.202	4.031	3.217	2.561
	P_2	9.126	3.763	3.189	2.322
	P_3	9.003	3.625	3.090	2.146
	O	6.789	3.310	2.913	2.616
	W	9.875	3.300	2.886	2.511
	M	10.786	4.411	3.900	2.602
5	P_1	9.801	3.280	3.010	2.330
	P_2	9.712	3.033	2.971	1.900
	P_3	8.791	3.168	2.811	2.176
	O	5.896	3.181	2.612	2.166
	W	8.442	2.901	2.077	1.896
	M	8.917	3.610	2.301	1.903
10	P_1	7.880	3.100	2.019	1.719
	P_2	8.721	2.911	1.891	1.778
	P_3	8.693	2.883	1.709	1.457

O = OLS
W = WLS
M = ML
P_1 = PL(1,1)
P_2 = PL(1,2)
P_3 = PL(2,2)

Table 4.9 Estimated variances of $\hat{\beta}$ for the OLS, WLS, ML, and PL (with different prior assumptions) estimators when the errors are Cauchy distributed with $\sigma_i = (0.5x_i+1)/3$, all i (all variances X100).

m	K→	4	6	8	10
2	O	7.088	6.127	6.795	5.322
	W	10.017	11.122	9.333	8.176
	M	12.602	12.787	9.712	9.011
	P_1	14.341	9.333	8.606	7.511
	P_2	15.132	7.255	8.510	5.037
	P_3	14.767	9.009	8.008	6.100
3	O	6.360	2.918	2.813	2.771
	W	10.910	3.717	3.439	3.292
	M	12.311	5.076	5.180	4.892
	P_1	11.765	5.008	4.777	3.321
	P_2	11.033	5.132	4.630	3.278
	P_3	11.100	5.160	4.339	3.371
4	O	7.881	4.117	3.817	2.176
	W	9.776	3.515	3.700	2.396
	M	9.890	4.133	2.311	3.101
	P_1	9.965	4.136	3.233	2.127
	P_2	8.765	3.986	3.681	2.039
	P_3	9.110	4.100	3.116	2.291
5	O	6.298	4.001	3.211	2.101
	W	9.438	3.796	3.011	2.289
	M	9.969	4.110	3.721	3.006
	P_1	9.822	3.717	2.988	1.917
	P_2	9.932	3.505	2.990	1.827
	P_3	8.876	3.127	2.716	1.886
10	O	4.218	3.022	2.008	1.961
	W	7.922	2.815	2.133	1.677
	M	8.011	3.147	2.900	1.812
	P_1	8.036	2.912	2.200	1.555
	P_2	7.911	3.007	2.103	1.605
	P_3	8.123	2.765	2.117	1.588

O = OLS
W = WLS
M = ML
P_1 = PL(1,1)
P_2 = PL(1,2)
P_3 = PL(2,2)

Table 4.10 Empirical biases of $\hat{\beta}$ for OLS, WLS, ML, and PL (with different prior assumptions) estimators when errors are normally distributed (biases are in percent).

		$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
m	K=→	4	6	8	10	4	6	8	10	4	6	8	10
2	O	0.901	-390	0.104	-0.105	1.495	-0.334	0.116	0.200	7.700	8.870	11.010	11.800
	W	-0.215	-0.791	-2.148	-0.594	-0.998	-1.206	-2.627	-0.800	-1.100	-0.850	-1.670	-0.600
	M	-1.128	-1.172	-2.612	-0.126	-1.482	-1.523	-3.426	-0.744	-1.690	-1.163	-2.891	-0.505
	P_1	0.207	-0.985	-0.757	0.154	-0.676	0.805	-0.976	0.968	0.913	-0.896	-0.364	0.711
	P_2	-0.665	0.552	-1.364	0.519	0.393	0.540	0.725	0.575	-0.808	0.473	-0.986	-0.476
	P_3	0.211	0.183	1.044	0.728	0.102	0.125	-0.061	0.042	0.317	0.318	0.007	-0.396
3	O	-0.281	-0.078	1.099	0.423	-0.566	-0.309	1.215	0.288	2.140	0.730	9.800	10.180
	W	-0.348	-0.496	1.151	-0.092	0.160	-0.314	1.877	-0.491	0.510	0.080	1.550	-0.500
	M	0.339	0.129	1.889	-0.244	1.070	1.208	1.935	0.837	-1.020	0.982	1.773	-0.701
	P_1	-0.121	-0.155	-0.582	0.265	0.089	0.273	-0.105	0.095	0.317	-0.261	-0.239	0.619
	P_2	0.427	0.027	0.160	-0.234	0.189	-0.715	-2.231	-1.205	-0.232	-0.099	-0.917	-0.380
	P_3	-0.160	-0.304	-0.789	-0.073	0.685	-0.282	-0.587	1.524	0.409	0.602	0.867	-0.271
4	O	-0.525	0.353	0.495	0.367	-1.263	0.326	0.870	0.396	-1.370	4.820	8.390	9.920
	W	-0.924	0.619	0.813	0.394	-1.060	0.602	0.869	0.300	0	0.270	0.400	0.364
	M	-0.973	0.620	0.980	0.239	-1.390	0.717	-0.191	-1.316	-0.091	0.389	0.376	-1.008
	P_1	-0.361	-0.616	-0.234	0.688	0.453	0.377	-0.664	-0.206	-0.201	0.235	-0.220	-0.966
	P_2	0.250	-0.458	-0.344	-0.371	-0.343	0.415	0.222	-0.813	-0.376	-0.187	-0.160	-0.182
	P_3	0.318	-0.342	-0.642	-0.038	-0.233	0.412	-0.099	0.652	0.108	0.239	-0.159	-0.298
5	O	1.374	-0.028	0.007	0.017	1.910	0.149	0.131	-0.100	1.490	2.370	1.570	8.130
	W	1.407	-0.002	-0.017	0.326	1.672	0.014	0.095	0.301	1.000	0.070	0.200	0.080
	M	1.283	-0.061	-0.128	0.271	-0.541	0.015	-0.863	0.855	0.800	-0.913	0.898	0.412
	P_1	-0.018	-0.147	-0.161	0.450	0.959	-0.567	0.587	0.125	0.432	-0.682	-0.769	-0.388
	P_2	-0.485	-0.054	-0.151	-0.539	0.040	-0.538	0.471	0.568	0.511	-0.080	-0.911	-0.161
	P_3	-0.385	0.213	0.026	0.782	1.416	-1.005	-0.633	-0.803	-0.401	0.789	0.622	0.303

Table 4.10 (continued)

		$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
m	K=→	4	6	8	10	4	6	8	10	4	6	8	10
10	O	0.221	−0.153	0.002	0.185	0.469	−0.267	−0.043	0.066	0.500	−0.250	0	0.010
	W	0.160	−0.282	0.041	0.306	0.462	−0.258	0.100	0.371	0.500	−0.390	0.050	0.170
	M	0.170	−0.266	0.033	0.325	0.065	0.072	−0.382	−0.487	0.579	−0.275	−0.185	−0.235
	P_1	0.314	−0.042	−0.093	0.070	−0.096	−0.021	−0.141	0.059	0.202	0.107	0.099	0.006
	P_2	−0.358	−0.008	−0.032	−0.028	−0.296	0.084	0.095	0.079	−0.189	−0.102	−0.173	−0.076
	P_3	0.024	0.128	0.042	−0.065	0.164	−0.009	0.002	−0.003	0.102	0.128	0.051	−0.102

Table 4.11 Empirical biases of $\hat{\beta}$ for OLS, WLS, ML, and PL (with different prior assumptions) estimators when errors are logistically distributed (biases are in percent).

m	K→	$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
		4	6	8	10	4	6	8	10	4	6	8	10
2	O	1.384	-0.390	0.156	2.278	1.407	-2.677	0.911	3.121	-1.777	-0.241	1.271	3.012
	W	3.994	-0.387	-1.281	1.206	-2.801	1.022	-1.099	1.606	-2.678	1.881	0.822	1.009
	M	3.900	-0.100	-2.001	1.269	0.710	30.011	-3.559	0.692	1.070	9.311	-4.788	-0.898
	P_1	1.198	-1.000	-0.986	-0.496	0.830	0.311	-0.824	0.211	-0.910	-0.239	-1.099	-0.999
	P_2	-1.176	0.501	0	-0.888	0.760	-0.084	0.321	-0.010	0.211	-0.103	0.600	0
	P_3	0.101	0	0.666	-0.285	0.351	0.312	0.729	0.266	-0.239	0.176	-0.289	0.017
3	O	2.035	-0.257	-0.925	-0.840	2.557	3.110	-1.000	1.001	1.993	0.506	-1.613	1.231
	W	2.137	-0.687	0.082	-1.775	2.609	-1.821	0.082	1.666	2.107	-1.072	-1.029	-0.987
	M	-0.489	0.333	1.100	8.200	-3.469	10.121	-3.688	-3.489	-2.133	-1.869	-3.900	-1.000
	P_1	0.500	0.206	0	-0.232	-0.829	0.711	0.071	0.007	0.717	0.950	0.102	0.120
	P_2	0.001	-0.485	0.286	-0.321	0.112	0.030	-0.288	0	0.681	-0.129	-0.191	-0.331
	P_3	-1.000	0.181	0	-0.081	0.096	-0.829	0.100	-0.188	1.710	-0.087	0	-0.024
4	O	-1.593	1.137	-0.517	-0.741	-0.389	0.633	-0.687	-0.667	-1.286	0.007	-0.999	1.000
	W	-1.836	0.633	-0.999	-0.637	-0.699	-0.002	-1.000	-0.554	-1.711	0.086	-1.001	1.006
	M	1.102	0.303	-0.388	0.424	7.670	9.131	10.971	1.797	8.181	7.312	0.876	0.690
	P_1	0.496	0.189	-1.866	-1.000	1.120	-0.069	0.101	0.891	0.965	-1.829	0.219	0.200
	P_2	-0.003	-0.689	0.300	0.722	-0.239	0.921	0.107	-0.100	-1.868	0.081	0.201	-0.120
	P_3	-0.139	0.865	0.391	-0.086	-0.879	-0.239	-0.099	0.121	0.015	-0.909	-1.009	0.606
5	O	-0.441	-0.014	0.150	-0.094	0.226	-1.007	-0.350	0.229	2.321	1.393	1.003	-1.967
	W	-0.335	-0.002	0.581	-0.471	0.009	0.012	-0.981	0.828	1.667	0.621	1.321	0.890
	M	-0.901	0.500	0.801	0.620	2.512	-2.824	2.711	1.251	2.712	-3.219	0.001	1.091
	P_1	0	0.191	0	0.210	1.121	-0.289	0.017	0.106	0.998	0.817	-0.179	0.098
	P_2	-0.081	0.606	-0.696	-0.323	-0.879	-0.621	-0.339	-0.158	-0.999	0.321	-0.089	-0.201
	P_3	0.863	-0.301	0.200	0.317	-0.088	0.891	0.019	0.100	-0.179	0.621	0.381	0.206

Table 4.11 (continued)

m	K⟶	$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
		4	6	8	10	4	6	8	10	4	6	8	10
10	O	-0.396	0.714	-0.101	0.701	-0.923	0.937	-1.283	-0.807	-1.015	0.920	-1.021	-1.065
	W	-0.610	0.489	-0.255	0.452	-1.211	1.110	-1.780	-0.411	-0.994	0.069	-0.621	-0.009
	M	0.289	0	0.601	0.612	0.511	-3.679	-2.388	-0.579	1.137	-0.632	-0.020	-1.788
	P_1	-0.188	0	0.037	-0.109	-0.219	0.117	0.079	0.110	-1.093	0.200	0.200	0.202
	P_2	0	0.090	0	-0.007	0.391	-0.183	0.092	-0.188	0.432	-0.088	0.101	-0.979
	P_3	0	0.088	0.007	0.081	-0.689	0	-1.000	0.031	0.621	0.011	-0.010	0.009

Table 4.12 Empirical biases of   for OLS, WLS, ML, and PL (with different prior assumptions) estimators when errors are Cauchy distributed (biases are in percent).

		$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
m	K⇒	4	6	8	10	4	6	8	10	4	6	8	10
2	O	-1.356	-3.750	-3.961	-3.751	-2.785	-2.780	-1.542	-1.767	2.035	-1.995	-0.777	-1.227
	W	-2.632	2.595	-0.900	1.511	-2.842	2.595	-1.514	0.407	1.994	-2.161	-3.117	-1.668
	M	-4.117	1.890	0.542	0.637	2.761	1.269	1.917	1.034	-1.775	-1.001	2.001	-3.811
	P_1	2.075	-2.410	-0.441	-1.130	-1.297	1.077	0.001	-0.110	0.200	0.876	0.688	0.172
	P_2	-1.028	-0.847	-0.846	0.681	0.660	-0.998	-0.076	0.034	-0.232	0.219	-0.422	0.982
	P_3	-0.299	-1.667	0.700	-0.471	0.982	0.607	-0.660	0.681	-0.321	0.201	-0.003	-0.617
3	O	1.732	-1.319	-6.163	-2.463	3.336	-1.736	-1.660	-1.320	0.961	-0.121	1.811	-1.489
	W	-5.449	-2.205	0.916	0.203	-1.364	-0.757	-0.164	-0.602	-1.007	1.812	-1.291	0.921
	M	-1.182	1.622	0.189	-0.313	-4.621	1.000	0.096	1.622	0.996	-1.912	2.112	-0.239
	P_1	-0.889	-0.486	-0.124	-0.596	0.960	0.298	0.032	-0.486	0.321	0.219	-0.210	-0.007
	P_2	0.240	0.284	0.762	-0.124	-0.167	0.063	-0.017	-0.202	0.729	-0.706	-0.323	0.012
	P_3	1.107	-1.714	0.127	-0.590	0.621	-1.009	0.003	-0.319	-1.000	0.608	0.317	-0.824
4	O	0.380	-1.231	-2.231	-2.371	-1.660	0.829	-6.803	-0.712	-2.001	4.212	-2.616	-1.809
	W	0.189	-0.311	-1.384	1.342	-0.866	0.091	0.548	-0.781	1.786	-1.003	2.131	-0.424
	M	-1.401	-1.551	0.149	0.767	-1.336	0.767	-0.261	0.918	0.071	-1.321	1.880	-1.000
	P_1	-3.327	-0.396	-0.794	0.768	-1.001	0.009	-0.093	0.118	-0.288	0.001	-0.701	0
	P_2	1.514	1.504	-0.870	1.432	-1.117	0.022	-0.099	-0.077	0.100	-0.179	-0.452	0.722
	P_3	-1.288	-2.151	0.330	0.904	1.007	0.001	0.692	-0.123	-0.687	-0.089	-0.612	-0.086
5	O	-1.966	3.825	-0.256	-3.292	1.528	-0.703	-1.765	-0.765	-2.789	2.381	2.912	-1.094
	W	2.270	2.345	-0.859	-1.879	0.909	0.783	-0.848	0.160	0.009	-1.021	-1.811	-0.471
	M	3.900	-0.205	0.231	1.606	1.900	-1.712	-1.006	1.200	0.012	-0.621	1.681	0.581
	P_1	-1.401	1.651	1.455	0.355	0.696	0.660	-0.913	-0.965	0.030	-1.000	-0.081	0.801
	P_2	0.700	1.390	-1.406	0.818	0.606	0.235	0.810	-0.623	-0.829	0	0.863	-0.696
	P_3	1.352	0.531	-1.168	-0.667	1.002	-0.198	-0.679	0.575	0.633	-0.911	-0.396	0.200

Table 4.12 (continued)

m	K→	$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
		4	6	8	10	4	6	8	10	4	6	8	10
10	O	−0.662	3.171	−0.775	−1.879	−0.896	−2.696	−1.035	−0.698	−2.141	−0.102	−0.660	0.101
	W	0.489	0.509	0.791	1.392	0.446	0.346	0.619	−0.034	3.121	1.200	1.771	−0.205
	M	−0.481	−0.667	0.855	0.283	0.360	0.478	−0.023	−1.717	−0.002	−0.107	−0.007	0.088
	P_1	0.065	0.327	0.758	0.006	0.680	−0.522	−1.001	0.652	0.239	0.811	−0.601	−0.634
	P_2	0.176	−0.296	0.503	−0.173	−0.210	−0.611	0.003	−0.332	−0.017	0.113	0	−0.006
	P_3	−0.044	−0.115	−0.931	−0.066	−0.029	−0.017	0.028	0.210	0.689	−1.001	−0.107	0.104

Table 4.13 Overview of minimum variance of $\hat{\beta}$ for the OLS, WLS, ML. and PL techniques with m, k, and σ combinations.

		$\sigma_i = 1$				$\sigma_i = (x_i+8)/9$				$\sigma_i = (0.5x_i+1)/3$			
	m K=→	4	6	8	10	4	6	8	10	4	6	8	10
Normal	2	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	PO	P	PW	0	OP	OP
	4	0	0	0	0	0	0	P	P	W	WP	P	OP
	5	0	0	0	0	OW	PW	WP	WP	W	W	WP	P
	10	OP	OP	OP	OP	MP	P	P	P	P	P	P	P
Logistic	2	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	P	0	0	0	OP	P	WOP	OWP	OWP	0
	4	OP	0	0	0	0	P	P	OP	WP	WP	PW	PW
	5	W	OW	0	P	0	OWP	WP	WMP	WP	WP	PW	P
	10	OP	P	P	OP	P	WP	P	P	P	P	PW	P
Cauchy	2	0	0	0	0	0	0	OP	OP	0	0	0	0
	3	0	W	0	OP	0	W	OW	OP	0	0	0	0
	4	0	OW	0	0	0	W	WP	WP	0	W	MP	OP
	5	0	OP	0	OM	0	WP	W	P	0	WP	P	P
	10	0	OP	M	P	OP	WP	P	P	OP	W	OP	P

0 = OLS W = WLS M = ML P = PL

CHAPTER 5

ADAPTIVE PROCEDURES IN THE ESTIMATION OF REGRESSION PARAMETERS

5.1 Introduction

In this chapter, we first describe the use of adaptive procedures for estimation. Then, in Section 5.2, we consider the existing techniques for testing variances and summarize comparative empirical studies of their power and robustness. Adaptive estimation procedures for the estimation of regression parameters after a preliminary test of variance equality are discussed in Section 5.3. In the last section empirical studies of the adaptive procedures are presented and discussed.

5.1.1 Preliminary Tests. When we are concerned with estimating the unknown parameters of behavioral and technical relations, there is uncertainty as to the appropriate model to be used. As a consequence, the researcher may begin with an initial set of specifications and then modify the models by testing the statistical significance of some or all of a class of hypotheses. This process makes the model, and thus the estimation procedure, dependent on the outcome of these tests of hypotheses which have been termed preliminary tests of significance. Thus most texts on statistical methods, including Snedecor and Cochran (1967), provide tests for nonconformity to model specification, e.g., tests for: non-normality of errors, outliers, homogeneity of variances, non-additivity, equality of means or correlation or regressions considered for pooling, etc., to be used as preliminary tests. If all the observations from an investigation are used both in making the preliminary tests and in providing the subsequent inferences of primary importance, the overall inference procedure is called conditionally specified inference.

The objective of conditionally specified inference procedures is to develop an objective stepwise inference methodology, with built-in preliminary tests of the critical uncertain elements, to minimize the consequences of such failures.

It is important to note that over 60 years ago R. A. Fisher (1920) suggested a conditionally specified inference procedure. He discussed what to do in a case where it is known that the example is from a population described by either a normal or a double exponential distribution. He suggested calculating the sample measure of kurtosis and recommended: If this is near 3, the MSE (mean square error) will be required (from which $\hat{\sigma} = \sqrt{MSE}$); if on the other hand, it approaches 6, its value for the double exponential curve, it may be that $\hat{\sigma}_1$ (based on the absolute deviation) is a more suitable measure of dispersion.

No theoretical studies of preliminary test procedures and their properties were available prior to 1944 (see Bancroft and Han, 1977). Since then, this class of statistical procedures has been studied, starting with Bancroft (1944) in the early 1940's, by Kitagawa (1963), Huntsberger (1955), Larson and Bancroft (1963), and Bancroft (1964), among others. The general approach has been to determine, usually for special cases, the properties of the resulting statistics in terms of their means and mean square errors. Scheffé (1959) also reviews research done to investigate the consequences of various types of failure in the model specification assumptions.

Bancroft and Han (1977) discussed estimation after preliminary testing for the random and fixed ANOVA models, using the never-pool test, the always-pool test and the sometimes-pool test.

Bancroft and Han (1977) also considered a different use of conditionally specified procedures in the fixed-effects ANOVA for multiple comparisons following a significant F-test. Usually when the treatments are declared to be different the investigator wishes to find the differences in specific treatment means. When multiple comparison procedures are used after the F-test of the treatment effect is significant, the F-test is essentially a preliminary test. The effect of the preliminary test should not be neglected. The effects of the preliminary test on error rates are studied by Berhardson (1975) and Smith (1974). Berhardson also reports the results of a Monte Carlo study. Ord and Leonard (1976) used an F-test for the equality of group means for fixed effects in the one-way ANOVA model as a preliminary test of significance and they provide a new procedure based upon the overall minimization of the mean square error.

5.2 Tests of Variance Equality

There are several test statistics available for testing the equality of population variances, such as those due to Bartlett, Cochran, and Hartley. However, as Box (1953) pointed out, these tests are sensitive to departures from normality in the underlying population being sampled and this sensitivity is due to the nonuse of within-group information. He suggested splitting the observations into subgroups and carrying out an analysis of variance on the logarithms of the subgroup sample variances.

Levene (1960) took a different approach by performing an analysis of variance on the transformed residuals from the mean. He suggested that tests based on either absolute residuals or squared residuals were satisfactory for power and robustness. Miller (1968) pointed out that Levene's test based on absolute residuals was not distribution-free when the underlying distributions were asymmetrical. He suggested that applying the jackknife technique to the sample variances before the analysis of variance would provide a robust test.

Layard (1973) modified the asymptotic chi-square test by using an estimator of population kurtosis different from that of Scheffé (1959). The results of his simulated sampling experiments indicate that the modified procedure performed as well as the jackknife test and compared favorably with Box's test.

He also suggested that a similar modification of Bartlett's test would improve the robustness of that test, although he did not include the test in his comparative study.

In Section 5.2.1 we consider two sample tests and then, in Section 5.2.2 we review $k(\geq 2)$ sample tests. Finally, in Section 5.2.3 we draw together the results of several empirical studies of the power and robustness of these different procedures.

5.2.1 Two-Sample Case - F-test. The most popular procedure for testing the equality of two variances is the variance ratio. First, we consider the situation where the observations are from normal populations. The location parameters may be unknown. The test statistic is

$$F = S_1^2/S_2^2 \tag{5.2.1}$$

where $$S_1^{\,2} = \Sigma(X_{i1}-\bar{X}_1)^2/(n_1-1),$$

$$S_2^{\,2} = \Sigma(X_{i2}-\bar{X}_2)^2/(n_2-1), \tag{5.2.2}$$

X_{i1} are the observations from the first sample,

X_{i2} are the observations from the second sample, and

n_1 and n_2 are the sample sizes. Also, $\bar{X}_1 = \Sigma X_{i1}/n_1$ and $\bar{X}_2 = \Sigma X_{i2}/n_2$ are the sample means. The procedure is called the two-sample F-test.

The usual hypothesis of interest states that the two variances are equal. Thus the hypothesis may be written as

(i) H_0: $\sigma_1^{\,2} = \sigma_2^{\,2}$ (ii) H_0: $\sigma_1^{\,2} \geqq \sigma_2^{\,2}$

H_1: $\sigma_1^{\,2} \neq \sigma_2^{\,2}$ H_1: $\sigma_1^{\,2} < \sigma_2^{\,2}$

or (iii) H_0: $\sigma_1^{\,2} \leqq \sigma_2^{\,2}$

H_1: $\sigma_1^{\,2} > \sigma_2^{\,2}$

The test statistic (5.2.1) is the appropriate statistic to use in testing all three hypotheses.

Both large and small values of F_{n_1-1,n_2-1} contradict H_0 in Case (i). Hence the critical region usually selected is

$$F_{n_1-1,n_2-1} < F_{\alpha/2;n_1-1,n_2-1} \text{ and } F_{n_1-1,n_2-1} > F_{1-\alpha/2;n_1-1,n_2-1} \tag{5.2.3}$$

In Case (ii), H_0 is rejected when $S_1^{\,2}$ is smaller to $S_2^{\,2}$. Thus the critical region is

$$F_{n_1-1,n_2-1} < F_{\alpha;n_1-1,n_2-1} \;. \qquad (5.2.4)$$

Finally, the critical region for situation (iii) is

$$F_{n_1-1,n_2-1} > F_{1-\alpha;n_1-1,n_2-1} \;. \qquad (5.2.5)$$

The F-test is extremely non-robust. Its actual significance level under the null hypothesis is much smaller than indicated for short-tailed distributions (uniform), and it gives too many significant results for long-tailed distributions such as the double exponential (see Miller, 1968).

5.2.2 The k-Sample Case. In this case, we consider the null hypothesis H_0: $\sigma_1^2=\ldots=\sigma_k^2$ against the general alternative that the variances are not all equal. There are several test procedures currently available for testing this null hypothesis, which we now describe.

Bartlett's Test

This procedure is an adaptation of a k-sample test constructed by Neyman. The test statistic is

$$B = (n-k)\log S^2 - \sum_{i=1}^{k}(n_i-1)\log S_i^2 \qquad (5.2.6)$$

where $$S^2 = \sum_{i=1}^{k}(n_i-1)S_i^2/(n-k) \qquad (5.2.7)$$

$$S_i^2 = \sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2/(n_i-1) \qquad (5.2.8)$$

and $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$ is the sample mean of the i^{th} sample.

The test was developed by Neyman using the likelihood ratio procedure. Consider the observations X_{ij}, $i=1,\ldots,k;j=1,2,\ldots,n_i$ where $X_{ij} \sim N(\mu_i,\sigma_i^2)$. The density function of X_{ij} is

$$P\{X_{ij}\} = (2\pi\sigma_i^2)^{-1}e^{-(X_{ij}-\mu_i)^2/2\sigma_i^2} \quad . \tag{5.2.9}$$

Then the likelihood function is

$$L = \frac{1}{(2\pi)^{n/2}\cdot\sigma_1^{n_1}\ldots\sigma_k^{n_k}} \exp\left[-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{X_{ij}-\mu_i}{\sigma_i}\right)^2\right] \tag{5.2.10}$$

where $\sum_{i=1}^{k} n_i = n$.

Let $L(\Omega)$ be the maximum of L over Ω, the entire parameter space of both the null and the alternative hypotheses. The maximum likelihood estimators are found to be

$$\hat{\mu}_i = \sum_{j=1}^{n_i} X_{ij}/n_i = \bar{X}_i; \quad \hat{\sigma}_i^2 = \sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2/n.$$

Substituting $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ into L we obtain

$$L(\Omega) = e^{-n/2}/(2\pi)^{n/2}\hat{\sigma}_1^{n_1}\ldots\hat{\sigma}_k^{n_k} \quad . \tag{5.2.11}$$

To compute $L(\omega)$, the maximum of L over ω, the parameter space under the null hypothesis, the maximum likelihood estimators used are

$$\hat{\mu}_i = \sum_{j=1}^{n_i} X_{ij}/n_i = \bar{X}_i; \; \hat{\sigma}^2 = \sum_{i=1}^{k} n_i\hat{\sigma}_i^2/n \quad . \tag{5.2.12}$$

Then,

$$L(\omega) = e^{-n/2}/(2\pi)^{n/2} \left\{ \sum_{i=1}^{k} n_i \hat{\sigma}_i^{\,2}/n \right\}^{n/2} . \qquad (5.2.13)$$

Substituting these into the likelihood ratio $\lambda = L(\omega)/L(\Omega)$ we obtain

$$\lambda = \prod_{i=1}^{k} \hat{\sigma}_i^{\,n_i}/\hat{\sigma}^n .$$

Using Wilk's result that $-2 \log \lambda$ has a $\chi^2_{(r-s)}$ distribution saymptotically, where r and s are the numbers of parameters in Ω and ω, we see that

$$-2 \log \lambda = - \sum_{i=1}^{k} n_i \log \frac{\hat{\sigma}_i^{\,2}}{\hat{\sigma}^2} \qquad (5.2.14)$$

is approximately distributed as $\chi^2_{(k-1)}$, under the null hypothesis,

$$H_0\colon \sigma_1^{\,2} = \ldots = \sigma_k^{\,2}.$$

Bartlett (1937) modified Neyman's result by using unbiased estimators for the variances and proposed the statistic

$$B = - \sum_{i=1}^{k} (n_i - 1) \log \frac{S_i^{\,2}}{S^2} . \qquad (5.2.15)$$

It has been found that a closer approximation to the distribution can be obtained by using the statistic B/C, where

$$C = 1 + \frac{1}{3(k-1)} \left[\sum_{i=1}^{k} \frac{1}{(n_i - 1)} - \frac{1}{\sum_{i=1}^{k} (n_i - 1)} \right] . \qquad (5.2.16)$$

In practice, usually B is computed first, then compared to the critical point $\chi^2_{\alpha,(k-1)}$. For $n_i \geq 2$ and $k \geq 2$, $1 \leq C \leq 5/3$. If it is possible that computing C could make a difference in the decision as to accept or reject H_0, then C is computed. Otherwise, it serves no purpose to compute C.

Modified Bartlett's Test

Bartlett's test is not accurate when the observations are not normally distributed. If we denote γ_2 as a measure of kurtosis, i.e.,

$$\gamma_2 = \sigma^{-4}E[(X-\mu)^4] - 3, \tag{5.2.17}$$

then Box (1953) points out that under H_0, $B \xrightarrow{L} [1+(\gamma_2/2)]\chi^2_{k-1}$ where $\xrightarrow{L}$ denotes convergence in distribution. If $\gamma_2 > 0$, then assuming that B is approximately χ^2 distributed gives too many significant results, while if $\gamma_2 < 0$, too few significant results are obtained. The modified statistic $B/[1+\hat{\gamma}_2/2]$, can provide a robust test. Several variants of this test have been proposed, cf. Scheffé (1959).

Hartley's Test

This procedure was developed by Hartley (1950) to provide a quick check of the homogeneity of k variances prior to conducting an analysis of variance. As in Bartlett's test, normality of the distributions is required. A further requirement is equality of sample sizes. The test is thus somewhat restricted, however, it is easily applied. The statistic is

$$F_{max} = S^2_{max}/S^2_{min} \tag{5.2.18}$$

where $$S^2_{max} = \max\{S_1{}^2,\ldots,S_k{}^2\}$$

$$S^2_{min} = \min\{S_1{}^2,\ldots,S_k{}^2\} \; .$$

The percentage points are calculated approximately using

$$F_{max}(\alpha) = \exp\{\omega_k(\alpha)\ \sqrt{\mathrm{Var}\ U}\ \} \tag{5.2.19}$$

where $U = \log S^2$

$\omega_k(\alpha)$ = $100\alpha\%$ points of the "range" ω in the independent samples of size k

$$\mathrm{Var}\ U \cong 2/n-1. \tag{5.2.20}$$

For small values of k and n, particularly for $n \leq 4$ and $k=2$ Hartley suggests that the values for $F_{max}(\alpha)$ should be adjusted using

$$F_{max}(\alpha) \triangleq F_{max}(\alpha)(1+q_n q_k) \tag{5.2.21}$$

where q_n and q_k are fitted to the exact values of k and n. However, when k=2, the two-sample F-test can be used so the value of the above approximation is questionable.

Significance points for this test are also given in David (1952) and Pearson and Hartley (1970, p. 202).

Cochran's Test

This test is somewhat similar to Hartley's test in its application. Again, for Cochran's test, the accuracy depends on the observations being normally distributed and the samples must be of the same size. The statistic is

$$C = S^2_{max} / \sum_{i=1}^{k} S_i^{\ 2} \qquad (5.2.22)$$

where $S^2_{max} = \max\{S_1^{\ 2}, \ldots, S_k^{\ 2}\}$.

This test is particularly snesitive to the case where all the variances $\sigma_k^{\ 2}$ are expected to be equal, except for one variance which could be large. Percentage points for the distribution of C are given in Dixon and Massey (1969, p. 536) or Pearson and Hartley (1970, p. 203); critical values not available can be obtained by quadrature from the approximation to the distribution function given by Cochran (1941, pp. 47-52).

Box's Test

This test was developed to counter the lack of robustness of other procedures with regard to the normality assumption. This procedure is usually called Box's test, as all sources indicate Box is responsible for the full development of the procedure, which involves performing an analysis of variance on the logarithms of the sample variances, although Box (1953) states was originally suggested by Bartlett and Kendall (1946).

The test was developed using the likelihood ratio procedure. Consider there are n_i observations in the i^{th} sample, and they are divided into t_i subgroups. Each subgroup contains m_{ij} observations and $\Sigma m_{ij} = n_i$. Let $y_{ij} = \log S^2_{ij}$, where S^2_{ij} represents the unbiased estimator of $\sigma_i^{\ 2}$ and is computed from m_{ij} observations.

If the null hypothesis is true, then

$$F = \frac{\Sigma t_i(\bar{y}_i-\bar{y})^2/(k-1)}{\Sigma\Sigma(y_{ij}-\bar{y}_i)^2/\Sigma(t_i-1)} \tag{5.2.23}$$

is approximately distributed as F with (k-1) and (Σt_i-k) d.f.

Cadwell's Test

For equal sample sizes, Cadwell (1953) proposed the test statistic

$$CW = \frac{\max(r_1,r_2,\ldots,r_k)}{\min(r_1,r_2,\ldots,r_k)} \tag{5.2.24}$$

where r_i is the range of the sample from the i^{th} population. Percentage points are given by Leslie and Brown (1965).

Modified Chi-Square Test

Layard (1973) suggested a chi-square test statistic

$$S' = \Sigma(n_i-1)\left(\log S_i^2 - \frac{\Sigma(n_i-1)\log S_i^2}{\Sigma(n_i-1)} \right)^2 \Bigg/ \left(2 + \left(1+\frac{1}{n}\right)\hat{\gamma}_2 \right) . \tag{5.2.25}$$

In estimating the amount of kurtosis (γ_2) of a distribution, Layard (1973) found that the resultant bias from using the weighted average of sample kurtoses can be reduced by using

$$\hat{\gamma}_2 = \frac{(\Sigma n_i)\Sigma\Sigma(X_{ij}-\bar{X}_i)^4}{(\Sigma\Sigma(X_{ij}-\bar{X}_i)^2)^2} - 3 \tag{5.2.26}$$

for non-normal populations. Under the null hypothesis, S' is asymptotically chi-square distributed with (k-1) degrees of freedom.

Jackknife Test

Miller (1968) proposed a procedure based on the jackknife technique to test H_0 in the two-sample case.

Layard (1973) proposed a generalization of Miller's (1968) test for the k-sample case.

Let

$$\bar{X}_{i(j)} = \frac{1}{(n_i-1)} \sum_{t \neq j} X_{it} \tag{5.2.27}$$

$$S_{i(j)} = \frac{1}{(n_i-2)} \sum_{t \neq j} (X_{it}-\bar{X}_{i(j)})^2 \tag{5.2.28}$$

and

$$y_{ij} = n_i \log S_i^2 - (n_i-1) \log S_{i(j)}^2 . \tag{5.2.29}$$

The jackknife test statistic

$$J = \frac{\Sigma n_i(\bar{y}_i-\bar{y})^2/(k-1)}{\Sigma\Sigma(y_{ij}-\bar{y}_i)^2/(\Sigma n_i-k)} \tag{5.2.30}$$

is distributed approximately as F with degrees of freedom (k-1) and (Σn_i-k) under the null hypothesis.

Modified Levene Test

Levene (1960) proposed a statistic for equal sample sizes which was subsequently generalized to unequal sample sizes. The statistic was obtained from a one-way ANOVA between groups, where each observation has been replaced by its absolute deviation from its group mean.

The modification is based on results in Miller (1968). That is, let $W_{ij} = |X_{ij}-M_i|$, where M_i is the i^{th} sample median instead of the sample mean as originally proposed by Levene (1960).

The analysis of variance test performed on the W_{ij} is distributed approximately as F with degrees of freedom (k-1) and $(\Sigma n_i - k)$ when all the population variances are equal.

5.2.3 Empirical Studies. Several comparative studies have been performed on the power and robustness of the different tests described in Section 5.2.2. The various studies and the tests compared are summarized in Table 5.1.

The findings of these studies are now summarized.

Miller (1968) showed that the F-test is extremely non-robust. Further, he found that jackknife test is reasonably robust and similar in power, and is more powerful than Box's test. However, Miller's study also found that the observed significance levels under the null hypothesis for the jackknife test are more sensitive to distributional form for large sample sizes. Finally, it was found that Levene's test is quite robust, but it lags far behind the jackknife test in power.

Gartside (1972) found that Bartlett's statistic was very powerful in all the experiments which he considered in his study. He suggested that if a short-cut test is needed then either Hartley's or Caldwell's statistic would be appropriate, as would the modified Bartlett's test. His empirical studies also showed that the modified Bartlett's test provide good power in all cases considered.

Layard (1973) showed that modified χ^2 and jackknife tests appear to be the best choices if a reasonably robust procedure is wanted. These tests are more powerful than Box's grouping test and perform similarly to Bartlett's test in the normal case.

Brown and Forsythe (1974) showed that the equality of variances in long-tailed distributions can best be tested by an alternative formulation of Levene's test statistic. These statistics use more robust estimators of central location in place of the mean. They are compared with unmodified Levene's statistic, a jackknife procedure and a χ^2 test suggested by Layard which are all found to be less robust under non-normality.

Keselman, Games, and Clinch (1979) compared six procedures which convert test of variance homogeneity into tests for mean equality for independent groups. They considered the samples of unequal sizes obtained from non-normal population. They found that Miller's jackknife procedure did not perform well when the sample sizes were unequal. They also recommended that for normal populations, Bartlett's test will be safe and more powerful.

5.2.4 Conclusion. From the above empirical studies, our findings are as follows:

(i) The F-test is extremely non-robust.

(ii) The modified χ^2 and jackknife tests are better and more powerful than Box's test but perform just like Bartlett's test in normal cases.

(iii) Jackknife procedures are not suitable for unequal sample sizes.

(iv) Modified Bartlett's test is robust and has good power properties. Also, it is easy to use.

On the basis of these conclusions, it was decided to use the modified Bartlett's test as a preliminary test of variance equality.

5.3 Adaptive Procedures for Estimation

In this section, we will be concerned with procedures involving the use of preliminary tests to decide whether to use a common variances (OLS) or different variances (WLS). For this purpose we consider both the modified Bartlett's test (MBT) and the determinants of variance $(\hat{\beta}_{WLS})$ and variance $(\hat{\beta}_{OLS})$ as preliminary test criteria.

In the first case our criteria will be as follows:

$$\hat{\underset{\sim}{\beta}}_A = \begin{cases} \hat{\underset{\sim}{\beta}}_{OLS} & \text{if MBT} < C; \\ \\ \hat{\underset{\sim}{\beta}}_{WLS} & \text{if MBT} \geq C; \end{cases} \tag{5.3.1}$$

here $\hat{\underset{\sim}{\beta}}_A$ represents $\hat{\underset{\sim}{\beta}}$ for the adaptive estimator after using a preliminary test. Also, C is an appropriate percentage points of the χ^2 distribution with (k-1) degrees of freedom. We also consider use of the generalized variances as follows:

$$\hat{\underset{\sim}{\beta}}_A = \begin{cases} \hat{\underset{\sim}{\beta}}_{OLS} & \text{if } |\mathrm{Var}(\hat{\beta}_{WLS})| \geq |\mathrm{Var}(\hat{\beta}_{OLS})|; \\ \\ \hat{\underset{\sim}{\beta}}_{WLS} & \text{if } |\mathrm{Var}(\hat{\beta}_{WLS})| < |\mathrm{Var}(\hat{\beta}_{OLS})|. \end{cases} \tag{5.3.2}$$

Now we will present the proof of unbiasedness of the adaptive estimator when distribution of $\underline{\epsilon}$ is symmetric about $\underline{0}$.

<u>Theorem 5.3.1</u>

Let the adaptive estimator, $\hat{\underset{\sim}{\beta}}_A$, based upon test statistic T, be

$$\text{if } T < t_o \quad \text{use OLS} \quad (\hat{\underset{\sim}{\beta}}_A = \hat{\underset{\sim}{\beta}}_{OLS}); \tag{5.3.3}$$

$$\text{if } T \geq t_o \quad \text{use WLS} \quad (\hat{\underset{\sim}{\beta}}_A = \hat{\underset{\sim}{\beta}}_{WLS}). \tag{5.3.4}$$

Then, when $\underline{\epsilon}$ is symmetric about $\underline{0}$, $E(\hat{\underset{\sim}{\beta}}_A) = \underset{\sim}{\beta}$.

Also let T be an even function of $\underline{\epsilon}$.

Proof: Consider

$$E(\hat{\underset{\sim}{\beta}}_A) = E(\hat{\underset{\sim}{\beta}}_{OLS}|T<t_o)P(T<t_o) + E(\hat{\underset{\sim}{\beta}}_{WLS}|T\geq t_o)P(T\geq t_o) \qquad (5.3.5)$$

Substituting for $\underline{Y}$ we obtain

$$E(\hat{\underset{\sim}{\beta}}_{OLS}|T<t_o) = \underline{\beta} + (\underline{X}'\underline{X})^{-1}X'E(\underline{\epsilon}|T<t_o). \qquad (5.3.6)$$

Since $\underline{\epsilon}$ is an odd function of $\underline{\epsilon}$ (trivially) and T is an even function of $\underline{\epsilon}$, it follows that, when the distribution of $\underline{\epsilon}$ is symmetric about the origin

$$E(\underline{\epsilon}|T<t_o) = E(-\underline{\epsilon}|T<t_o) \therefore E(\underline{\epsilon}|T<t_o) = 0. \qquad (5.3.7)$$

Thus, (5.3.7) implies that (5.3.6) will be

$$E(\hat{\underset{\sim}{\beta}}_{OLS}|T<t_o) = \underline{\beta}. \qquad (5.3.8)$$

Similarly,

$$E(\hat{\underset{\sim}{\beta}}_{WLS}|T\geq t_o) = \underset{\sim}{\beta} + E\{(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{\epsilon}|T\geq t_o\} \qquad (5.3.9)$$

$$= \underset{\sim}{\beta} + E\{G(\underline{\epsilon})|T\geq t_o\} \qquad (5.3.10)$$

where $G(\underline{\epsilon})$ is the function inside the expectation operator in (5.3.9). Arranging in the same way as before, it follows that $G(\underline{\epsilon})$ is an odd function of $\underline{\epsilon}$ and T is an even function of $\underline{\epsilon}$, so that

$$E\{G(\underline{\epsilon})|T\geq t_o\} = E\{G(-\underline{\epsilon})|T\geq t_o\}$$

$$= - E\{G(\underline{\epsilon})|T\geq t_o\}.$$

Hence, using the symmetry of the distribution of $\underline{\epsilon}$ about zero,

$$E\{G(\underset{\sim}{\epsilon})|T\geq t_o\} = \underset{\sim}{0}. \tag{5.3.11}$$

Thus, using (5.3.11), (5.3.10) becomes

$$E(\hat{\underset{\sim}{\beta}}_{WLS}|T\geq t_o) = \underset{\sim}{\beta}\ . \tag{5.3.12}$$

Finally, substituting (5.3.8) and (5.3.12) into (5.3.5), we have

$$E(\hat{\underset{\sim}{\beta}}_A) = \underset{\sim}{\beta}\ . \tag{5.3.13}$$

In Section 5.3.1, we present empirical results on the performance of the MBT and determinant criteria as preliminary tests. Conclusions about empirical studies are presented in the last section.

5.3.1 Empirical Results. The empirical findings for the estimators described in Section 5.3 are presented here as follows:

In Table 5.2, the empirical results after using MBT as a preliminary test criteria are presented and the empirical results after using the generalized variances as a preliminary test criteria are presented in Table 5.3.

5.3.2 Conclusions. In Table 5.2, after using the modified Bartlett's test as a preliminary test, our findings about adaptive estimators are as follows:

(i) The adaptive estimators have variances which are comparable with the better of OLS and WLS. That is, the adaptive estimators perform almost as well as if the correct model was known.

(ii) The "best" significance level varies with the variance structure; α=0.10 seems better for σ=1 and $\sigma=(X_i+8)/9$ whereas α=0.01 is generally better for the remaining case. Indeed, an inappropriate choice of α may largely nullify the benefits of the preliminary test.

In Table 5.3, the adaptive estimators based upon the generalized variance criterion perform similarly to those based upon the modified Bartlett's test, although for the "best" choice of α; the variances are slightly smaller. Nevertheless, since the generalized variance criterion avoids the need for choosing α, it is felt that it provides better estimators overall.

A final problem with any adaptive procedure is the estimation of the variance of the estimator. A natural rule to use in practics is "use the variance expression for the selected estimator," although we may suspect that this is downwards biased. From Table 5.3, we find an average downward bias is the variance of about 6% when m=10 and about 12% when m=5. For these particular models, this does not seem to be a major concern, but clearly further work is required to establish better estimators for the variance.

Table 5.1 Comparative studies of the tests of equality of variances by different authors.

Author(s)	B	MB	$M\chi^2$	JK	BOX	F	LEVENE	C	H	CW	$BF_{\bar{X}}$	BF_m
Layard (1973)	*		*	*	*							
Gartside (1972)	*	*						*	*	*		
Miller (1968)				*		*	*					
Geng, Wang, and Miller (1979)	*	*	*	*	*		*					
Brown and Forsythe (1974)			*	*		*	*				*	*
Keselman, Games, and Clinch (1979)			*								*	*

B = Bartlett
MB = Modified Bartlett
$M\chi^2$ = Modified χ^2
JK = Jackknife
C = Cochran

$BF_{\bar{X}}$ = Brown and Forsythe (trimmed mean)
BF_m = Brown and Forsythe (absolute deivation from median)
H = Hartley
CW = Cadwell

Table 5.2 Empirical results of the adaptive estimator of $\underline{\hat{\beta}}_A$ after using MBT (Modified Bartlett Test) as a preliminary test.

α	0.20	0.10	0.05	0.01	WLS	OLS
			(σ_i=1; m=10; k=4)			
$E(\hat{\beta})$	0.98969	0.98935	0.98838	0.98800	0.99440	0.98795
$Var(\hat{\beta})$	0.00243	0.00239	0.00250	0.00248	0.00288	0.00249
	OLS=157 WLS=43	OLS=178 WLS=22	O=192 W=8	O=198 W=2		
			(σ_i=1; m=10; k=6)			
$E(\hat{\beta})$	0.98927	0.98907	0.98888	0.99075	0.98692	0.99085
$Var(\hat{\beta})$	0.00167	0.00165	0.00165	0.00166	0.00181	0.00166
	O=164 W=36	O=186 W=14	O=192 W=8	O=199 W=1		
			(σ_i=1; m=10; k=8)			
$E(\hat{\beta})$	1.00466	1.00361	1.00514	1.00420	1.00296	1.00420
$Var(\hat{\beta})$	0.00156	0.00153	0.00154	0.00155	0.00175	0.00155
	O=157 W=43	O=182 W=18	O=195 W=5	O=200 W=0		
			(σ_i=1; m=10; k=10)			
$E(\hat{\beta})$	0.99479	0.99432	0.99691	0.99693	0.99650	0.99707
$Var(\hat{\beta})$	0.00136	0.00133	0.00131	0.00131	0.00147	0.00132
	O=161 W=39	O=182 W=18	O=194 W=6	O=198 W=2		
			($\sigma_i=(\frac{1}{2}X_i+1)/3$; m=10; k=4)			
$E(\hat{\beta})$	0.99815	0.99753	0.99896	0.99901	0.99890	0.98256
$Var(\hat{\beta})$	0.00382	0.00379	0.00379	0.00375	0.00382	0.00393
	O=3 W=197	O=11 W=189	O=35 W=165	O=86 W=114		

Table 5.2 (Continued)

α	0.20	0.10	0.05	0.01	WLS	OLS
			$(\sigma_i = (\frac{1}{2} X_i+1)/3;$ m=10, k=6)			
$E(\hat{\beta})$	1.00235	0.99781	0.98147	0.98456	1.00237	0.97978
$Var(\hat{\beta})$	0.00269	0.00275	0.00271	0.00263	0.00270	0.00340
	O=3 W=197	O=11 W=189	O=22 W=178	O=76 W=124		
			$(\sigma_i = (\frac{1}{2} X_i+1)/3;$ m=10; k=8)			
$E(\hat{\beta})$	0.99688	0.99312	0.99712	0.99910	0.99703	1.00743
$Var(\hat{\beta})$	0.00210	0.00209	0.00207	0.00213	0.00210	0.00290
	O=3 W=197	O=9 W=191	O=20 W=180	O=78 W=122		
			$(\sigma_i = (\frac{X}{2}+1)/3;$ m=10; k=10)			
$E(\hat{\beta})$	0.98782	0.98698	0.99391	0.99901	0.99885	0.98888
$Var(\hat{\beta})$	0.00186	0.00188	0.00186	0.00181	0.00186	0.00260
	O=15 W=185	O=12 W=188	O=29 W=171	O=82 W=118		
			$(\sigma_i = (X_i+8)/9;$ m=10; k=4)			
$E(\hat{\beta})$	0.99998	0.99379	0.98730	0.99828	0.99702	0.99787
$Var(\hat{\beta})$	0.00580	0.00578	0.00577	0.00589	0.00579	0.00590
	O=66 W=134	O=112 W=88	O=138 W=62	O=182 W=18		
			$(\sigma_i = (X_i+8)/9;$ m=10; k=6)			
$E(\hat{\beta})$	1.00041	0.98365	0.99362	0.99078	1.00238	0.97909
$Var(\hat{\beta})$	0.00401	0.00397	0.00423	0.00399	0.00411	0.00486
	O=62 W=138	O=111 W=89	O=140 W=60	O=183 W=17		

Table 5.2 (Continued)

	0.20	0.10	0.05	0.01	WLS	OLS
			($\sigma_i = (X_i+8)/9$; m=10; k=8)			
$E(\hat{\beta})$	1.00035	0.99817	1.00585	1.01036	0.99520	1.00775
$Var(\hat{\beta})$	0.00370	0.00377	0.00401	0.00437	0.00373	0.00442
	O=67 W=133	O=104 W=96	O=135 W=65	O=192 W=8		
			($\sigma_i = (X_i+8)/9$; m=10; k=10)			
$E(\hat{\beta})$	0.99299	0.98969	0.99898	0.98966	0.99681	0.98653
$Var(\hat{\beta})$	0.00307	0.00303	0.00316	0.00330	0.00309	0.00334
	O=77 W=123	O=110 W=90	O=149 W=51	O=183 W=17		

Table 5.3 Empirical results of the adaptive estimators of $\hat{\underline{\beta}}_A$ after using generalized variance as a preliminary test.

	[m=10; σ_i=1]				[m=10; σ_i=(0.5x_i+1)/3]			
k→	4	6	8	10	4	6	8	10
$E(\hat{\beta})$	0.99678	1.00099	0.99661	0.99853	0.99890	1.00237	0.99701	0.99885
$V(\hat{\beta})$	0.00258	0.00172	0.00171	0.00137	0.00424	0.00269	0.00209	0.00187
$V(\hat{\beta}\hat{\beta})$	0.00227	0.00154	0.00164	0.00125	0.00381	0.00265	0.00213	0.00186
Ratio	1.13656	1.11688	1.04268	1.09600	1.11286	1.01509	0.98122	1.00538
	OLS=94 WLS=106	OLS=92 WLS=108	OLS=99 WLS=101	OLS=95 WLS=105	OLS=1 WLS=199	OLS=0 WLS=200	OLS=0 WLS=200	OLS=0 WLS=200

	[m=10; σ_i=(x_i+8)/9]				[m=5; σ_i=1]			
k→	4	6	8	10	4	6	8	10
$E(\hat{\beta})$	0.99718	1.00216	0.99520	0.99681	1.00308	0.99779	0.99810	1.00030
$V(\hat{\beta})$	0.00579	0.00422	0.00373	0.00309	0.00469	0.00346	0.00390	0.00305
$V(\hat{\beta}\hat{\beta})$	0.00550	0.00401	0.00364	0.00305	0.00425	0.00355	0.00340	0.00276
Ratio	1.05273	1.05237	1.02473	1.01311	1.10353	0.97465	1.14706	1.10507
	OLS=2 WLS=198	OLS=0 WLS=200	OLS=0 WLS=200	OLS=0 WLS=200	OLS=85 WLS=115	OLS=92 WLS=108	OLS=103 WLS=97	OLS=92 WLS=108

	[m=5; σ_i = (0.5x_i+1)/3]				[m=5; σ_i = (x_i+8)/9]			
K→	4	6	8	10	4	6	8	10
$E(\hat{\beta})$	1.00485	0.99660	1.00065	0.99710	1.00509	0.99534	0.99686	0.99963
$V(\hat{\beta})$	0.00712	0.00633	0.00547	0.00527	0.01066	0.00887	0.01071	0.00890
$V(\hat{\beta}\hat{\beta})$	0.00601	0.00588	0.00490	0.00484	0.00915	0.00890	0.00839	0.00717
Ratio	1.18469	1.07653	1.11633	1.08884	1.16503	0.99663	1.27652	1.24128
	OLS=3 WLS=197	OLS=2 WLS=198	OLS=0 WLS=200	OLS=0 WLS=200	OLS=5 WLS=195	OLS=5 WLS=195	OLS=2 WLS=198	OLS=4 WLS=196

$V(\hat{\beta})$ = variance estimated from 200 samples

$V(\hat{\beta}\hat{\beta})$ = variance on the basis of individual sample (using the OLS or WLS variance approximation)

Ratio = $V(\hat{\beta})/V(\hat{\beta}\hat{\beta})$

CHAPTER 6

USE OF ADAPTIVE PROCEDURES IN THE ESTIMATION OF REGRESSION PARAMETERS BY GROUPING COMMON VARIANCES

6.1 Introduction

We have already presented the concept and purpose behind using adaptive procedures in Chapter 5. On the same lines we develop other adaptive procedures in this chapter. Suppose we have k mean square estimates S_i^2 ($i=1,2,\ldots,k$), each based on m degrees of freedom, of the variances σ_i^2 of k normal populations. The overall equality of variances may be tested by one of the test statistics provided in Section 5.2. But statistical significance on these overall tests leave open the question as to which variances differ. So we now seek to identify those variances which may be pooled to further improve the estimators for the regression model.

We will use a variant of Fisher's Least Significant Difference test (FSD) as a preliminary test for grouping common variances, that is, we perform an F test for each pair of variances and compare the ratios to a modified set of percentage points.

A brief review of multiple comparisons is given in Section 6.2 and the FSD procedure is outlined in Section 6.3. The empirical results are then presented and summarized in Section 6.4.

6.2 Multiple Comparison Methods

Numerous procedures are available for performance of pairwise multiple comparisons among the observed treatment means from designed experiments (see O'Neill and Wetherill, 1971; Miller, 1966, 1977; Carmer and Swanson, 1973; and Stoline, 1981). The methods discussed in the above-mentioned papers are Least Significant Difference method (LSD), the Fisher Significant Difference method (FSD), Tukey's T-method, Scheffé's S method, Newman-Keuls method, Student-Newman-Keuls method, Duncan's multiple range method, the Tukey (1953) method, the Bonferroni method (c.f. Miller, 1966, p. 8), Dunn-Šidak method and Tukey-Kramer (TK) method.

An experimenter is often interested in making inferences about variances instead of, or in addition to, inferences about means. There are many procedures which exist for testing the hypothesis of homogeneity of variances (see Section 5.2) of k populations, but these procedures do not determine which variances differ from one another. David (1956) proposed a multiple range test for variances on the basis of Duncan's (1955) philosophy. Levy (1975) proposed three different range tests for variances based on the approach of Newman (1939) and Keuls (1952). These three tests are based on F-max statistic, Cochran's statistic and a normalizing log transformation of the sample variance. Hartley (1955) gave a sequential F test for multiple comparisons of mean squares. This procedure consists in comparing the ratios S_i^2/S_j^2, taken in descending order of magnitude, with appropriate percentage points, until a non-significant ratio is achieved.

6.2.1 Properties of Different Procedures. The commonly used multiple comparison tests suggested by different authors, Fisher's test (FSD) based on Bonferroni's inequality and Scheffé's test for all contrasts are very conservative. The user of Scheffé's test pays a penalty of added conservatism for its versality; this test should never be used if one is interested only in comparing pairs of means. Tukey's studentized range test is somewhat conservative. Progressively less conservative tests are Tukey's X procedure, the Newman-Keuls test, and Duncan's multiple range test. Least conservative of all is the LSD test. Non-conservative tests provide poor control over the error rate per experiment (or experimentwise). Conservative tests, on the other hand, may limit the error rate per comparison to unnecessary low values, and tend to have low power unless the sample size is large. The relation between error rates and sample sizes for range tests in multiple comparisons was studied by Harter (1957).

Seegar (1968) recommends the FSD method very strongly, partly because it is much more flexible in use. O'Neill and Weitherill (1971) recommend FSD, LSD, T and S-methods for a small or moderate number of contrasts. Conover and Iman (1981) suggest Scheffé's, Tukey's, and FSD methods for their simplicity and flexibility. Carmer and Swanson (1973) also like the FSD test, which has good control of type I errors, more power than the Newman-Keuls, Studentized range or S-methods, and easy to use.

On the basis of the above discussion and conclusion, it was decided to use the FSD test as a preliminary test for multiple comparisons of variance equality.

6.3 A Multiple Comparisons Test for the Error Variances

Based upon the survey in Section 6.2, we use the variance form of Fisher's LSD test in order to decide whether to group variances. Thus, given k sample variances $S_1^2,\ldots,S_k^2$, we form the test statistics

$$F_{ij} = \max(S_i^{\ 2}, S_j^{\ 2})/\min(S_i^{\ 2}, S_j^{\ 2}) \quad i=1,2,\ldots,k;\ j=1,2,\ldots,k \tag{6.3.1}$$

(F_{ii}=1, trivially, for all i).

Given an appopriate percentage point C_{ij}, if

$$F_{ij} < C_{ij} \tag{6.3.2}$$

we do not reject

$$H_0\colon \sigma_i^2 = \sigma_j^2 \ . \tag{6.3.3}$$

If $F_{ij} > C_{ij}$, H_0 is rejected. This is coded as

$$\begin{aligned} \delta_{ij} &= 1, \text{ if } H_0 \text{ is not rejected} \\ \delta_{ij} &= 0, \text{ if } H_0 \text{ is rejected.} \end{aligned} \tag{6.3.4}$$

Given these test results we use the estimator

$$\tilde{\sigma}_i^2 = \sum_{j=1}^{k} \delta_{ij}(m_j-1)S_j^{\ 2} \Big/ \sum_{j=1}^{k} \delta_{ij}(m_j-1) \tag{6.3.5}$$

with the convention that δ_{ii}=1 for all i. That is, if the F test does not reject H_0 in (6.3.3), then $S_j^{\ 2}$ is used in $\tilde{\sigma}_i^2$. When each $S_i^{\ 2}$ has (m-1) degrees of freedom, all the C_{ij} values will be equal. The appropriate values of C for different values of m and k and different overall significance levels α^* are summarized in Table 6.1.

The C values were calculated using the Bonferroni inequality with $\alpha^* = \alpha\Big/\binom{k}{2}$, where α is actual level of significance and $\binom{k}{2}$ is the number of combinations of k populations taken two at a time. The method of interpolation used in the construction of the table was as follows:

Let α_1, α_2 be values for which percentage points are tabulated

(F_{α_1} and F_{α_2} say) and let $\alpha^*(\alpha_1<\alpha^*<\alpha_2)$ be the value of interest. Then the interpolated value for F at level α^* is

$$F_{\alpha^*} = (1-A)F_{\alpha_2} - AF_{\alpha_1} \qquad (6.3.6)$$

where

$$1 - A = (\alpha_2)^{-0.5} - (\alpha^*)^{-0.5}/(\alpha_2)^{-0.5} - (\alpha_1)^{-0.5} \quad . \qquad (6.3.7)$$

Various powers of α were tried in (6.3.7) and evaluated by using the existing tables of percentage points, the value -0.5 was found to perform best. Although more accurate interpolation formulae could be constructed, the results in Table 6.1 are sufficiently close for present use.

6.4 Conclusion

The results of the simulation study for multiple comparisons are presented in Table 6.2. For the preliminary tests are utilized a significance level of α=0.10; some results were also obtained for α=0.25, but appeared to differ only slightly. The number of replicates were m=5 and m=10 for each of the three patterns of variances considered.

The conclusions of the study are as follows:

(i) As expected, the estimators show no systematic bias.

(ii) Although there is considerable variation, it appears that the preliminary testing procedures may yield sizeable reduction in the variances.

(iii) On the basis of the summary in Table 6.3, the FSD testing procedure produces estimates with the smallest variances. Thus the FSD procedure supercedes all the other procedures (OLS, WLS, Modified Bartlett and Generalized Variance) and can be recommended as a superior procedure.

Table 6.1 Upper percentage points of the variance ratio in a set of k variances each based on m-1 degrees of freedom (normal variation assumed).

m \ k	2	4	6	8	10
		(a) $\alpha = 0.25$			
2	3.0000	23.9648	62.5226	113.4235	181.7498
3	2.2798	10.4954	19.8170	29.8265	40.7813
4	2.0000	7.1291	11.5202	15.7197	19.8337
5	1.8528	5.6979	8.4056	10.8049	12.9880
6	1.7622	4.9237	6.8426	8.4479	9.8307
7	1.7010	4.4433	5.9198	7.0984	8.0696
8	1.6369	4.1176	5.3161	6.2350	6.9642
9	1.6236	3.8830	4.8925	5.6394	6.2123
10	1.5975	3.7061	4.5799	5.2055	5.6707
		(b) $\alpha = 0.10$			
2	9.0000	62.4383	153.5109	318.0945	523.4800
3	5.4624	20.3609	37.0820	56.9587	78.3277
4	4.3248	12.0675	18.8739	25.4550	31.9384
5	3.7797	8.9355	12.7530	15.9925	19.0067
10	2.9245	5.0580	6.0137	6.5736	7.0188
		(c) $\alpha = 0.05$			
2	19.0000	122.4878	345.1644	633.8663	999.4001
3	9.5521	32.7490	60.8694	91.090	129.2000
4	6.9443	17.6351	27.1805	36.3676	48.0510
5	5.7861	12.3538	17.8673	21.3955	26.9207
10	4.1028	6.3058	7.1617	7.7909	8.7399

Table 6.2 A comparative study of the mean and variances of $\hat{\beta}$ before and after using the FSD test as a preliminary test.

$k=\rightarrow$	4	6	8	10	4	6	8	10	4	6	8	10
	($m=10$; $\alpha=0.10$; $\sigma_i=1$)				($m=10$; $\alpha=0.10$; $\sigma_i=(0.5x_i+1)/3$)				($m=10$; $\alpha=0.10$; $\sigma_i=(x_i+8)/9$)			
$E(\hat{\beta}_A)$	1.00200	0.99977	1.00090	0.99659	0.99932	1.00186	1.00267	0.99830	1.00118	1.00089	1.00258	0.99647
$E(\hat{\beta})$	0.99440	0.98692	1.00296	0.99650	0.99890	1.00237	0.99703	0.99885	0.99702	1.00238	0.99520	0.99681
$Var(\hat{\beta}_A)$	0.00258	0.00144	0.00150	0.00123	0.00341	0.00225	0.00183	0.00134	0.00552	0.00326	0.00319	0.00233
$Var(\hat{\beta})$	0.00288	0.00181	0.00175	0.00147	0.00382	0.00270	0.00210	0.00186	0.00579	0.00411	0.00373	0.00309
	($m=5$; $\alpha=0.10$; $\sigma_i=1$)				($m=5$; $\alpha=0.10$; $\sigma_i=(0.5x_i+1)/3$)				($m=5$; $\alpha=0.10$; $\sigma_i=(x_i+8)/9$)			
$E(\hat{\beta}_A)$	1.00123	0.99576	1.00096	0.99746	1.00134	0.99447	0.99641	0.99451	1.00185	0.99298	0.99858	0.99423
$E(\hat{\beta})$	1.01407	0.99998	0.99983	1.00326	1.01000	1.00070	1.00200	1.00831	1.00472	1.00014	1.00095	1.00301
$Var(\hat{\beta}_A)$	0.00439	0.00372	0.00413	0.00262	0.00640	0.00560	0.00541	0.00411	0.00947	0.00758	0.00894	0.00699
$Var(\hat{\beta})$	0.00527	0.00412	0.00420	0.00325	0.00710	0.00632	0.00547	0.00526	0.01066	0.00887	0.01071	0.00889

α = level of significance for percentage points; $\hat{\beta}_A$ = mean of 200 simulated values of $\hat{\beta}$ after using preliminary test; $Var(\hat{\beta}_A)$ = average variance of 200 simulated values of $\hat{\beta}$ after using preliminary test; $E(\hat{\beta})$ and $Var(\hat{\beta})$ are the actual mean and variance of $\hat{\beta}$ on the basis of WLS.

Table 6.3 A comparative study of the different estimators for $\hat{\beta}$ on the basis of Tables 5.2, 5.3, and 6.2 (the procedure with the smallest variance is recorded in each case).

	m = 10			m = 5		
k	$\sigma_i=1$	$\sigma_i = (0.5x_i+1)/3$	$\sigma_i = (x_i+8)/9$	$\sigma_i=1$	$\sigma_i = (0.5x_i+1)/3$	$\sigma_i = (x_i+8)/9$
4	M	F	F	F	F	F
6	F	F	F	O	F	F
8	F	F	F	O	F	F
10	F	F	F	F	F	F

Tests considered:

O = OLS
W = WLS
M = Modified Bartlett (α=0.10)
F = FSD (α=0.10)
G = Generalized Variance

CHAPTER 7

ESTIMATION FOR THE ERRORS-IN-VARIABLES MODEL

7.1 Introduction

Variables in a regression model may be masked by measurement errors which arise from different factors or hidden sources. Nevertheless, it may be possible to make inferences about the parameters relating to the regression of true variables. This problem has been examined extensively by many authors, e.g., Kendall and Stuart (1979), Cochran (1968), Mandansky (1959), Moran (1971), Sprent (1966, 1969), and Villegas (1961, 1964). These authors draw attention to the variety of, ad hoc methods of estimation available, including "grouping" methods and the use of instrumental variables, cumulants, and components of variance.

The first detailed application of general methods of estimation in the two variable linear errors-in-variables problem is that of Lindley (1947). He resolved many of the earlier uncertainties and anomalies in demonstrating the breakdown of the maximum likelihood method, reflecting in the unidentifiability of the parameters and the inconsistency of the estimators for the linear structural, and linear functional, models respectively. But this was specifically for the unreplicated case. Not too much is written on the replicated case.

We develop the results with the help of maximum likelihood in the next section for the replicated case. As Barnett (1970) points out, "In principle there is no reason why the maximum likelihood method should not be used in the replicated case, apart from the computational problems of unravelling the sometimes awakward ML equations, but the author is not aware of any published results on this."

7.2 Maximum Likelihood Estimators (Equal Variances Case)

Consider the linear case of two variables

$$X_i = \eta_i + \delta_i \tag{7.1}$$

$$Y_{ij} = \xi_i + \epsilon_{ij} \tag{7.2}$$

$$\xi_i = \beta_0 + \beta_1\eta_i \tag{7.3}$$

$$Y_{ij} = \beta_0+\beta_1X_i+(\epsilon_{ij}-\beta_1\delta_i),\ i=1,2,\ldots,k;\ j=1,2,\ldots,m. \tag{7.4}$$

where,

$$\epsilon_{ij} \sim IN(0,\sigma^2_\epsilon);\ \delta_i \sim IN(0,\sigma^2_\delta);\ Cov(\epsilon_{ij},\delta_i) = 0 \tag{7.5}$$

X_i and Y_{ij} are observed values; η_i and ξ_i are true values and δ_i and ϵ_{ij} are errors in observations which are mutually and serially independent.

The likelihood function for the sample observations is

$$L = \prod_{i=1}^{k} \prod_{j=1}^{m} \frac{1}{\sigma_\epsilon\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2_\epsilon}(Y_{ij}-\beta_0-\beta_1\eta_i)^2\right] \prod_{i=1}^{k} \frac{1}{\sigma_\delta\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2_\delta}(X_i-\eta_i)^2\right] \tag{7.6}$$

so the log likelihood is

$$\ell = \text{const.} - \frac{1}{2}\,km\,\log(\sigma^2_\epsilon) - \frac{1}{2}\,k\,\log(\sigma^2_\delta) - \frac{1}{2}\,\frac{\sum_{i=1}^{k}\sum_{j=1}^{m}(Y_{ij}-\beta_0-\beta_1\eta_i)^2}{\sigma^2_\epsilon} - \frac{1}{2}\sum_{i=1}^{k}\frac{(X_i-\eta_i)^2}{\sigma^2_\delta} \tag{7.7}$$

We take the derivative of Eq. 7.7 w.r. to β_0, β_1, η_i, σ^2_δ, and δ^2_ϵ and equate to zero. This yields the following equations:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{\eta} = \bar{Y} - \hat{\beta}_1\bar{X} \tag{7.8}$$

$$\hat{\beta}_1 = \frac{\Sigma\Sigma(Y_{ij}-\bar{Y})(\hat{\eta}_i-\bar{\eta})}{\Sigma(\hat{\eta}_i-\bar{\eta})^2} = \frac{\Sigma(\bar{Y}_{i.}-\bar{Y})(\hat{\eta}_i-\bar{\eta})}{\Sigma(\hat{\eta}_i-\bar{\eta})^2} \tag{7.9}$$

$$\hat{\sigma}^2_\delta = \frac{1}{k}\,\Sigma(X_i-\hat{\eta}_i)^2 = \frac{1}{k}\,\Sigma[(X_i-\bar{X})-(\hat{\eta}_i-\bar{\eta})]^2 \tag{7.10}$$

$$\hat{\sigma}^2_\epsilon = \frac{1}{km}\,\Sigma\Sigma(Y_{ij}-\hat{\beta}_0-\hat{\beta}_1\hat{\eta}_i)^2 = \frac{1}{km}\,\Sigma\Sigma[Y_{ij}-\bar{Y}-\hat{\beta}_1(\hat{\eta}_i-\bar{\eta})]^2 \tag{7.11}$$

and

$$\left(\frac{1}{\hat{\sigma}^2_\delta} + \frac{\hat{\beta}_1^{\,2}}{\hat{\sigma}^2_\epsilon}\right)(\hat{\eta}_i-\bar{\eta}) = \frac{\hat{\beta}_1}{\hat{\sigma}^2_\epsilon}(\bar{Y}_{i.}-\bar{Y}) + \frac{(X_i-\bar{X})}{\hat{\sigma}^2_\delta}\,. \tag{7.12}$$

Substituting (7.8) in Eq. 7.12, we get

$$(\bar{Y}_{i.}-\hat{\beta}_0-\hat{\beta}_1\hat{\eta}_i) = \frac{\hat{\sigma}^2_\epsilon}{\hat{\beta}_1\hat{\sigma}^2_\delta}(X_i-\hat{\eta}_i). \tag{7.13}$$

Substituting Eq. 7.13 in Eq. 7.11, and using 7.10

$$\hat{\sigma}^2_\epsilon = \frac{1}{km}\,\Sigma\Sigma(Y_{ij}-\bar{Y}_{i.}+\bar{Y}_{i.}-\hat{\beta}_0-\hat{\beta}_1\hat{\eta}_i)^2$$

$$\hat{\sigma}^2_\epsilon = \frac{1}{km}\,\Sigma\Sigma[(Y_{ij}-\bar{Y}_{i.}) + \frac{\hat{\sigma}^2_\epsilon}{\hat{\sigma}^2_\delta\hat{\beta}_1}(X_i-\hat{\eta}_i)]^2$$

which yields

$$\hat{\sigma}^2_\epsilon = \mathrm{Tyy}/km\left[1 - \frac{1}{\lambda\hat{\beta}_1^{\,2}}\right], \tag{7.14}$$

where $$Tyy = \Sigma\Sigma(Y_{ij}-\bar{Y}_{i.})^2 \tag{7.15}$$

and $$\lambda = \hat{\sigma}^2_\delta/\hat{\sigma}^2_\epsilon \ . \tag{7.16}$$

From Eq. 7.12, substituting the value of $(\hat{\eta}_i-\bar{\eta})$ in Eq. 7.10

$$k\hat{\sigma}^2_\delta\left(\frac{1}{\hat{\sigma}^2_\delta}+\frac{\hat{\beta}_1^{\,2}}{\hat{\sigma}^2_\epsilon}\right)^2 = \frac{\hat{\beta}_1^{\,2}}{\hat{\sigma}^4_\epsilon}[\hat{\beta}_1^{\,2}S_{XX}+S_{\bar{Y}\bar{Y}}-2\hat{\beta}_1S_{X\bar{Y}}]$$

$$\frac{k\hat{\sigma}^2_\epsilon}{\lambda}(1+\hat{\beta}_1^{\,2}\lambda)^2 = \hat{\beta}_1^{\,2}[\hat{\beta}_1^{\,2}S_{XX}+S_{\bar{Y}\bar{Y}}-2\hat{\beta}_1S_{X\bar{Y}}]$$

or $$\hat{\sigma}^2_\epsilon = \frac{\hat{\beta}_1^{\,2}\lambda}{k(1+\hat{\beta}_1^{\,2}\lambda)^2}[\hat{\beta}_1^{\,2}S_{XX}+S_{\bar{Y}\bar{Y}}-2\hat{\beta}_1S_{X\bar{Y}}] \tag{7.17}$$

where $\Sigma(X_i-\bar{X})^2 = S_{XX}$; $\Sigma(\bar{Y}_{i.}-\bar{Y})^2 = S_{\bar{Y}\bar{Y}}$ and $\Sigma(X_i-\bar{X})(\bar{Y}_{i.}-\bar{Y}) = S_{X\bar{Y}}$.

(7.18)

From Eq. 7.12, substituting the value of $(\hat{\eta}_i-\bar{\eta})$ in Eq. 7.9

$$\hat{\beta}_1 = \frac{\Sigma[\bar{Y}_{i.}-\bar{Y}]\left[\frac{\hat{\beta}_1}{\hat{\sigma}^2_\epsilon}(\bar{Y}_{i.}-\bar{Y})+\frac{X_i-\bar{X}}{\hat{\sigma}^2_\delta}\right]\left[\frac{1}{\hat{\sigma}^2_\delta}+\frac{\hat{\beta}_1^{\,2}}{\hat{\sigma}^2_\epsilon}\right]}{\Sigma\left[\frac{\hat{\beta}_1}{\hat{\sigma}^2_\epsilon}(\bar{Y}_{i.}-\bar{Y})+\frac{X_i-\bar{X}}{\hat{\sigma}^2_\delta}\right]^2} \ .$$

Substituting Eq. 7.18 and Eq. 7.16, we obtain

$$\lambda = \frac{S_{X\bar{Y}}-\hat{\beta}_1S_{XX}}{\hat{\beta}_1(\hat{\beta}_1S_{X\bar{Y}}-S_{\bar{Y}\bar{Y}})} \ . \tag{7.19}$$

Substituting Eq. 7.14 in Eq. 7.17

$$\frac{1}{m} T_{YY} = \frac{(\lambda\hat{\beta}_1^{\,2}-1)(\hat{\beta}_1^{\,2} S_{XX}+S_{\overline{Y}\overline{Y}} - 2\hat{\beta}_1 S_{X\overline{Y}})}{(1+\hat{\beta}_1^{\,2}\lambda)^2} \quad . \tag{7.20}$$

Substituting Eq. 7.19 in Eq. 7.20, we get

$$\frac{1}{m} T_{YY} = \frac{(S_{\overline{Y}\overline{Y}} - \hat{\beta}_1^{\,2} S_{XX})(\hat{\beta}_1 S_{X\overline{Y}} - S_{\overline{Y}\overline{Y}})}{(\hat{\beta}_1^{\,2} S_{XX}+S_{\overline{Y}\overline{Y}} - 2\hat{\beta}_1 S_{XY})} \quad .$$

Eq. 7.20 may be written as

$$(mS_{XX}S_{X\overline{Y}})\hat{\beta}_1^{\,3} + (S_{XX}T_{YY}-mS_{XX}S_{\overline{Y}\overline{Y}})\hat{\beta}_1^{\,2}-(2S_{X\overline{Y}}T_{YY}+mS_{X\overline{Y}}S_{\overline{Y}\overline{Y}})\hat{\beta}_1$$

$$+(mS^2_{\overline{Y}\overline{Y}} + S_{\overline{Y}\overline{Y}}T_{YY}) = 0 \tag{7.21}$$

Once the solution to 7.21 is obtained, we compute λ from 7.19 and hence $\hat{\sigma}^2_{\epsilon}$ from 7.17, $\hat{\sigma}^2_{\delta}$ from 7.16.

7.3 Maximum Likelihood Estimators (Different Variances Case)

Equations 7.1-7.4 are the same, save that we now assume

$$\epsilon_{ij} \sim N(0,\sigma^2_{\epsilon i}); \; \delta_i \sim N(0,\sigma^2_{\delta}); \; \mathrm{Cov}(\epsilon_{ij},\delta_i) = 0 \; . \tag{7.22}$$

Taking this log likelihood function

$$\ell = \text{const} - \frac{1}{2} m \log(\sigma^2_{\epsilon i}) - \frac{1}{2} k \log(\sigma^2_{\delta}) - \frac{1}{2}\sum_i \left[\sum_j \frac{(Y_{ij}-\beta_0-\beta_1\eta_i)^2}{\sigma^2_{\epsilon i}}\right]$$

$$- \frac{1}{2}\sum_i \frac{(X_i-\eta_i)}{\sigma^2_{\delta}} \tag{7.23}$$

Taking the derivative of Eq. 7.23 w.r.t. β_0, β_1, η_i, σ_δ^2, and $\sigma_{\epsilon i}^2$ and equate to zero. This yields the following equations

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{\eta} = \bar{Y} - \hat{\beta}_1\bar{X} \tag{7.24}$$

$$\hat{\beta}_1 = \frac{\Sigma\Sigma(Y_{ij}-\bar{Y})(\hat{\eta}_i-\bar{\eta})}{\Sigma(\hat{\eta}_i-\bar{\eta})^2} = \frac{\sum_i(\bar{Y}_{i.}-\bar{Y})(\hat{\eta}_i-\bar{\eta})}{\sum_i(\hat{\eta}_i-\bar{\eta})^2} \tag{7.25}$$

$$\hat{\sigma}_\delta^2 = \frac{1}{k}\sum_i(X_i-\hat{\eta}_i)^2 = \frac{1}{k}\Sigma[(X_i-\bar{X})-(\hat{\eta}_i-\bar{\eta})]^2 \tag{7.26}$$

$$\hat{\sigma}_{\epsilon i}^2 = \frac{1}{m}\sum_j(Y_{ij}-\hat{\beta}_0-\hat{\beta}_1\hat{\eta}_i)^2 = \frac{1}{m}\sum_j(Y_{ij}-\bar{Y}_{i.}+\bar{Y}_{i.}-\hat{\beta}_0-\hat{\beta}_1\hat{\eta}_i)^2 \tag{7.27}$$

and
$$\left(\frac{1}{\hat{\sigma}_\delta^2} + \frac{\hat{\beta}_1^{\,2}}{\hat{\sigma}_{\epsilon i}^2}\right)(\hat{\eta}_i-\bar{\eta}) = \frac{\hat{\beta}_1}{\hat{\sigma}_{\epsilon i}^2}(\bar{Y}_{i.}-\bar{Y}) + \frac{(X_i-\bar{X})}{\hat{\sigma}_\delta^2} \ . \tag{7.28}$$

After substituting Eq. 7.24 into Eq. 7.28, we get

$$(\bar{Y}_{i.}-\hat{\beta}_0-\hat{\beta}_1\hat{\eta}_i) = \frac{\hat{\sigma}_{\epsilon i}^2}{\hat{\beta}_1\hat{\sigma}_\delta^2}(X_i-\hat{\eta}_i) \ . \tag{7.29}$$

Substituting Eq. 7.28 into Eq. 7.25, we get

$$\hat{\beta}_1 = \frac{\sum_i[\hat{\beta}_1\hat{\sigma}_\delta^2(\bar{Y}_{i.}-\bar{Y})^2+\hat{\sigma}_{\epsilon i}^2(X_i-\bar{X})(\bar{Y}_{i.}-\bar{Y})][\hat{\sigma}_{\epsilon i}^2+\hat{\beta}_1^{\,2}\sigma_\delta^2]}{\sum_i[\hat{\beta}_1^{\,2}\hat{\sigma}_\delta^4(\bar{Y}_{i.}-\bar{Y})^2+\hat{\sigma}_{\epsilon i}^4(X_i-\bar{X})^2+2\hat{\beta}_1\hat{\sigma}_{\epsilon i}^2\hat{\sigma}_\delta^2(\bar{Y}_{i.}-\bar{Y})(X_i-\bar{X})]} \ . \tag{7.30}$$

Substituting Eq. 7.28 into Eq. 7.26, we obtain

$$\hat{\sigma}_\delta^2 = \frac{1}{k}\sum_i\frac{\hat{\beta}_1^{\,2}\sigma_\delta^4}{(\hat{\sigma}_{\epsilon i}^2+\hat{\beta}_1^{\,2}\hat{\sigma}_\delta^2)^2}[\hat{\beta}_1(X_i-\bar{X})-(\bar{Y}_{i.}-\bar{Y})]^2 \ . \tag{7.31}$$

Substituting Eqs. 7.29, 7.28, and 7.24 into Eq. 7.27

$$\hat{\sigma}^2_{\epsilon i} = \frac{T_{ii}}{m} + \frac{[\hat{\beta}_1(X_i-\bar{X})-(\bar{Y}_{i.}-\bar{Y})]^2}{\left[1+\frac{\hat{\beta}_1{}^2\hat{\sigma}^2_\delta}{\hat{\sigma}^2_{\epsilon i}}\right]^2} \tag{7.32}$$

where $T_{ii} = \sum_j (Y_{ij}-\bar{Y}_{i.})^2$. (7.33)

Further simplification of these results does not appear to be possible and we must find the ML estimators by iterative solution from Eqs. 7.30-7.32, starting with the results for the equal variances case and substituting into Eq. 7.32, then 7.30 and 7.31 in that order. The iterations continue until the solution converges.

7.4 Large Sample Properties of ML (Common Variance Case)

Eqs. 7.1-7.7 are the same. Taking the second derivatives of Eq. 7.7 w.r. to β_0, β_1, σ^2_δ, and σ^2_ϵ are as follows

$$\partial^2\ell/\partial\beta_0{}^2 = -mk/\hat{\sigma}^2_\epsilon \tag{7.34}$$

$$\partial^2\ell/\partial\beta_1{}^2 = -\sum_{ij}\sum(\hat{\eta}_i-\bar{\eta})^2/\hat{\sigma}^2_\epsilon \tag{7.35}$$

$$\partial^2\ell/\partial(\sigma^2_\delta)^2 = k/2\hat{\sigma}^4_\delta - \sum_i(X_i-\hat{\eta}_i)^2/\hat{\sigma}^6_\delta \tag{7.36}$$

$$\partial^2\ell/\partial(\sigma^2_\epsilon)^2 = km/2\hat{\sigma}^4_\epsilon - \sum_{ij}\sum[Y_{ij}-\hat{\beta}_0-\hat{\beta}_1(\hat{\eta}_i-\bar{\eta})]^2/\hat{\sigma}^6_\epsilon \tag{7.37}$$

and after simplification and taking expectations, these second derivatives will be equal to zero

$$E(\partial^2\ell/\partial\beta_0\partial\beta_1) = E(\partial^2\ell/\partial\beta_0\partial\sigma_\epsilon^2) = E(\partial^2\ell/\partial\beta_0\partial\sigma_\delta^2) = E(\partial^2\ell/\partial\beta_1\partial\sigma_\delta^2)$$

$$= E(\partial^2\ell/\partial\beta_1\partial\sigma_\epsilon^2) = E(\partial^2\ell/\partial\sigma^2\partial\sigma_\epsilon^2) = 0 \tag{7.38}$$

Taking the expectations of the Eqs. 7.34-7.37

$$-E[\partial^2\ell/\partial\beta_0{}^2] = mk/\hat{\sigma}_\epsilon^2 \tag{7.39}$$

$$-E[\partial^2\ell/\partial\beta_1{}^2] = mk(\sigma_x^2-\sigma_\delta^2)/\sigma_\epsilon^2 \tag{7.40}$$

where, we define σ_x^2 as $\lim\limits_{k\to\infty} \frac{1}{k} \sum\limits_i (X_i-\bar{X})^2$ (7.41)

$$-E[\partial^2\ell/\partial(\sigma_\delta^2)^2] = k/2\hat{\sigma}_\delta^4 \tag{7.42}$$

$$-E[\partial^2\ell/\partial(\hat{\sigma}_\epsilon^2)^2] = mk/2\hat{\sigma}_\epsilon^4 \tag{7.43}$$

$$\mathrm{Var}_{(ML)} = \begin{bmatrix} -E(\partial^2\ell/\partial\beta_0^2) & -E(\partial^2\ell/\partial\beta_0\partial\beta_1) & -E(\partial^2\ell/\partial\beta_0\partial\sigma_\epsilon^2) & -E(\partial^2\ell/\partial\beta_0\partial\sigma_\delta^2) \\ & -E(\partial^2\ell/\partial\beta_1{}^2) & -E(\partial^2\ell/\partial\beta_1\partial\sigma_\epsilon^2) & -E(\partial^2\ell/\partial\beta_1\partial\sigma_\delta^2) \\ & & -E(\partial^2\ell/\partial(\sigma_\epsilon^2)^2) & -E(\partial^2\ell/\partial\sigma_\epsilon^2\partial\sigma_\delta^2) \\ & & & -E(\partial^2\ell/\partial(\sigma_\delta^2)^2) \end{bmatrix}^{-1} \tag{7.44}$$

After substituting Eqs. 7.39-7.43 into Eq. 7.44

$$\mathrm{Var}_{(ML)} = \begin{bmatrix} mk/\hat{\sigma}_\epsilon^2 & 0 & 0 & 0 \\ 0 & mk(\sigma_x^2-\sigma_\delta^2)/\sigma_\epsilon^2 & 0 & 0 \\ 0 & 0 & mk/2\sigma_\epsilon^4 & 0 \\ 0 & 0 & 0 & k/2\sigma_\epsilon^4 \end{bmatrix}^{-1} \tag{7.45}$$

Hence

$$\left.\begin{aligned} Var(\hat{\beta}_0) &= \hat{\sigma}_\epsilon^2/mk \\ Var(\hat{\beta}_1) &= \hat{\sigma}_\epsilon^2/mk(\sigma_x^2-\sigma_\delta^2) \\ Var(\hat{\sigma}_\epsilon^2) &= 2\hat{\sigma}_\epsilon^4/mk \\ Var(\hat{\sigma}^2) &= 2\hat{\sigma}_\delta^4/k \end{aligned}\right\} \quad (7.46)$$

The estimates are consistent as $k \to \infty$.

7.5 Large Sample Properties of ML (Unequal Variances Case)

Eqs. 7.1-7.4 and Eqs. 7.22-7.23 are the same. Taking the second derivatives of 7.23 w.r. to β_0, β_1, σ_δ^2, and $\sigma_{\epsilon i}^2$.

$$\partial^2 \ell/\partial \beta_0{}^2 = -m \sum_i \frac{1}{\hat{\sigma}_{\epsilon i}^2} \quad (7.47)$$

$$\partial^2 \ell/\partial \beta_1{}^2 = -m \sum_i [(\hat{\eta}_i-\bar{\eta})^2/\hat{\sigma}_{\epsilon i}^2] \quad (7.48)$$

$$\partial^2 \ell/\partial (\sigma_\delta^2)^2 = k/2\sigma_\delta^4 - \sum_i [(X_i-\hat{\eta}_i)^2/\hat{\sigma}_\delta^6] \quad (7.49)$$

$$\partial^2 \ell/\partial (\sigma_{\epsilon i}^2)^2 = m/2\hat{\sigma}_{\epsilon i}^4 - \sum_j (Y_{ij}-\hat{\beta}_0-\hat{\beta}_1\hat{\eta}_i)^2/\hat{\sigma}_{\epsilon i}^6 \quad (7.50)$$

$$\partial^2 \ell/\partial \beta_0 \partial \beta_1 = -m \sum_i [(\hat{\eta}_i-\bar{\eta})/\hat{\sigma}_{\epsilon i}^2] \quad (7.51)$$

and after simplification or taking expectations, these derivatives are zero

$$E(\partial^2\ell/\partial\beta_0\partial\sigma_\delta^2) = E(\partial^2\ell/\partial\beta_1\partial\sigma_\delta^2) = E(\partial^2\ell/\partial\beta_0\partial\sigma_{\epsilon i}^2) = E(\partial^2\ell/\partial\beta_1\partial\sigma_{\epsilon i}^2)$$

$$= E(\partial^2\ell/\partial\sigma_\delta^2\partial\sigma_{\epsilon i}^2) = 0 \, . \quad (7.52)$$

Taking the expectation of Eqs. 7.47-7.51

$$-E(\partial^2\ell/\partial\beta_0{}^2) = \sum_i [m/\sigma_{\epsilon i}^2] \tag{7.53}$$

$$-E(\partial^2\ell/\partial\beta_1{}^2) = m\sum_i [(\hat{\sigma}_x^2-\hat{\sigma}_\delta^2)/\hat{\sigma}_{\epsilon i}^2] \tag{7.54}$$

$$-E(\partial^2\ell/\partial(\sigma_\delta^2)^2) = k/2\hat{\sigma}_\delta^2 \tag{7.55}$$

$$-E(\partial^2\ell/\partial(\sigma_{\epsilon i}^2)^2) = m/2\hat{\sigma}_{\epsilon i}^4 \tag{7.56}$$

$$-E(\partial^2\ell/\partial\beta_0\partial\beta_1] = m\sum_i (X_i-\bar{X})/\hat{\sigma}_{\epsilon i}^2) \tag{7.57}$$

$$\mathrm{Var}_{(ML)} = \begin{bmatrix} \sum_i (m/\hat{\sigma}_{\epsilon i}^2) & m\sum_i [(X_i-\bar{X})/\hat{\sigma}_{\epsilon i}^2] & 0 & 0 \\ m\sum_i [(X_i-\bar{X})/\hat{\sigma}_{\epsilon i}^2] & m\sum_i [(\hat{\sigma}_x^2-\sigma_\delta^2)/\hat{\sigma}_{\epsilon i}^2] & 0 & 0 \\ 0 & 0 & m/2\hat{\sigma}_{\epsilon i}^4 & \\ 0 & 0 & 0 & k/2\hat{\sigma}_\delta^2 \end{bmatrix}^{-1} . \tag{7.58}$$

Hence from Eq. 7.58

$$\left.\begin{aligned} \mathrm{Var}(\hat{\beta}_0) &= (D-B)/(AD-B^2) \\ \mathrm{Var}(\hat{\beta}_1) &= (A-B)/(AD-B^2) \\ \mathrm{Var}(\hat{\sigma}_{\epsilon i}^2) &= 2\hat{\sigma}_{\epsilon i}^4/m \\ \mathrm{Var}(\hat{\sigma}_\delta^2) &= 2\hat{\sigma}_\delta^2/k \end{aligned}\right\} \tag{7.59}$$

where $A = \sum_i (m/\hat{\sigma}_{\epsilon i}^2)$; $B = m\sum_i [(X_i-\bar{X})/\hat{\sigma}_{\epsilon i}^2]$

and $D = m\sum_i [(\hat{\sigma}_x^2-\hat{\sigma}_\delta^2)/\hat{\sigma}_{\epsilon i}^2]$.

Note: $\hat{\sigma}_{\epsilon i}^2$ is consistent if $m\to\infty$ and $\hat{\sigma}_\delta^2$ is consistent if $k\to\infty$.

7.6 Method of Least Squares

We now consider how the approach from LS regression analysis breaks down when applied to the estimation of β_0 and β_1 in Eq. 7.4, even if the errors δ_i and ϵ_{ij} are assumed to be mutually and serially independent with constant variances, and also to be independent of the true values η_i and ξ_i. The application of least squares to Eq. 7.4 to get estimates of β_0 and β_1 is not valid, since the factor $(\epsilon_{ij}-\beta_1\delta_i]$ in Eq. 7.4 is not independent of X_i. The covariance of X_i and $[\epsilon_{ij}-\beta_1\delta_i]$ is

$$\mathrm{Cov}[X_i, (\epsilon_{ij}-\beta_1\delta_i)] = -\beta_1 \mathrm{Var}(\delta_i) \text{ using Eq. 7.5.} \quad (7.60)$$

Since the covariance does not vanish, a dependence between error term and explanatory variables in Eq. 7.4.

Due to this dependence the application of LS to Eq. 7.4 would yield biased estimates of the β_0 and β_1. Furthermore, the bias will not dieappear as the sample size becomes infinitely large; so the LS estimates are inconsistent. The bias in the replicated case is the same as in the unreplicated case, as we now show.

<u>7.4.1 Inconsistency of L. S. Estimators</u>. The least square estimators of β_1 on the basis of J observations in k samples is

$$\beta_1^* = \frac{\sum\limits_{ij}\sum(X_i-\bar{X})(Y_{ij}-\bar{Y})}{\sum\limits_i(X_i-\bar{X})^2} = \frac{\sum\limits_i(X_i-\bar{X})(\bar{Y}_{i.}-\bar{Y})}{\sum\limits_i(X_i-\bar{X})^2} \quad . \quad (7.61)$$

Taking limits in probability as follows:

$$\underset{k\to\infty}{\mathrm{p\ lim}} \left[\frac{1}{k} \Sigma(X_i-\bar{X})(\bar{Y}_{i.}-\bar{Y})\right] = \mathrm{p\ lim} \frac{1}{k} \Sigma(\eta_i-\bar{\eta})(\beta_0+\beta_1\eta_i+\bar{\epsilon}_{i.})$$

$$= p\lim[0 + \frac{1}{k}\beta_1\Sigma(\eta_i-\bar{\eta})^2+\Sigma\bar{\epsilon}_{i.}(\eta_i-\bar{\eta})]$$

$$= \beta_1 \lim[\Sigma(\eta_i-\bar{\eta})^2/k] = \beta_1\sigma_\eta^2 \qquad (7.62)$$

$$p\lim_{k\to\infty} \frac{1}{k}\Sigma(X_i-\bar{X})^2 = p\lim \frac{1}{k}\Sigma[(\eta_i-\bar{\eta})+(\delta_i-\bar{\delta})]^2 = \sigma_\eta^2+\sigma_\delta^2 . \qquad (7.63)$$

Substituting Eq. 7.36-7.37 into Eq. 7.35, we obtain

$$p\lim_{k\to\infty} \beta^*_1 = \beta_1\sigma_\eta^2/(\sigma_\eta^2+\sigma_\delta^2)$$

or $$p\lim_{k\to\infty} \beta^*_1 = \beta_1(1-\sigma_\delta^2/\sigma_x^2) . \qquad (7.64)$$

Thus $p\lim_{k\to\infty} \beta^*_1 \neq \beta_1$, but is in fact an underestimate of β_1.

The asymptotic mean square errors for the ML and OLS estimators are as follows

$$MSE(ML) = \hat{\sigma}_\epsilon^2/Mk(\hat{\sigma}_x^2-\hat{\sigma}_\delta^2) \qquad (7.65)$$

and $$MSE(OLS) = (\hat{\beta}_1\hat{\sigma}_\delta^2/\hat{\sigma}_x^2)^2 + \hat{\sigma}_\epsilon^2/mk\hat{\sigma}_x^2 . \qquad (7.66)$$

Approximately, MSE(ML) may be greater than MSE(OLS) if

$$\hat{\sigma}_\epsilon^2/mk(\hat{\sigma}_x^2-\hat{\sigma}_\delta^2) > \hat{\beta}_1{}^2\hat{\sigma}_\delta^4/\sigma_x^4 + \hat{\sigma}_\epsilon^2/mk\hat{\sigma}_x^2$$

or $$\hat{\sigma}_\epsilon^2/mk(\hat{\sigma}_x^2-\hat{\sigma}_\delta^2) > \hat{\beta}_1{}^2\hat{\sigma}_\delta^2/\hat{\sigma}_x^2 \qquad (7.67)$$

i.e., if $|\hat{\beta}_1|$ is small

or if $\hat{\sigma}_\delta^2\ (1-\hat{\sigma}_\delta^2/\hat{\sigma}_x^2)$ is small i.e., errors-in-variables are relatively small or if mk is small.

Because of this we undertook an empirical study to compare the performance of the ML, OLS, and WLS estimators for both common and different variances. We will present the empirical study in the next section.

7.7 Empirical Results

The study presented in Tables 7.1 and 7.2 shows that: (i) The ML estimator of $\hat{\beta}$ has very little bias comparison to WLS and OLS but high MSE because of high variability in the errors in the X variables when small samples are used. (ii) When σ_δ^2 is small, then the OLS estimates have a small bias as expected. However, when σ_δ^2 is large, the MSE for OLS is still smaller than that for ML unless the number of replications is increased. (iii) When σ_δ^2 is large and large size samples with a large number of replicates are used, ML gives better estimates.

Table 7.1 A comparative study of ML, WLS and OLS for $\hat{\beta}$ in the "errors-in-variables" case when error variances are different.

	$[\sigma_{\epsilon i}=1;\ \sigma_{\delta}=1]$			$[\sigma_{\epsilon i}=(0.5x_i+1)/3;\ \sigma_{\delta}=1]$			$[\sigma_{\epsilon i}=(X_i+8)/9;\ \sigma_{\delta}=1]$		
	$(\bar{\hat{\beta}})$	$Var(\hat{\beta})$	$MSE(\hat{\beta})$	$(\bar{\hat{\beta}})$	$Var(\hat{\beta})$	$MSE(\hat{\beta})$	$(\bar{\hat{\beta}})$	$Var(\hat{\beta})$	$MSE(\hat{\beta})$
				(k=4; m=10)					
ML	1.00602	0.18831	0.18835	1.05137	0.22038	0.22302	1.08487	0.25744	0.26464
WLS	0.97730	0.02333	0.02385	0.97772	0.03132	0.03808	0.97599	0.02789	0.02847
OLS	0.90780	0.02157	0.03005	0.93773	0.02295	0.02683	0.91801	0.02353	0.02361
				[k=10; m=10]					
ML	0.95708	0.01181	0.01365	1.01262	0.01840	0.01856	1.05892	0.02898	0.03245
WLS	0.92724	0.00886	0.01415	0.91815	0.01326	0.01664	0.92558	0.01053	0.01607
OLS	0.92541	0.00883	0.01439	0.92488	0.00986	0.01550	0.92484	0.01044	0.01609
	$[\sigma_{\epsilon i}=1;\ \sigma_{\delta}=2]$			$[\sigma_{\epsilon i}=(0.5x_i+1)/3;\ \sigma_{\delta}=2]$			$[\sigma_{\epsilon i}=(X_i+8)/9;\ \sigma_{\delta}=2]$		
				(k=4; m=10)					
ML	0.96101	0.13651	0.13803	1.00333	0.13021	0.13022	1.07127	0.17011	0.17519
WLS	0.93121	0.09330	0.09806	0.93110	0.08992	0.09441	0.98121	0.10022	0.10054
OLS	0.93878	0.08518	0.08902	0.93839	0.08552	0.08924	0.96710	0.09307	0.09409
				[k=10; m=10]					
ML	0.98170	0.01911	0.01947	0.99354	0.02515	0.02519	1.00231	0.03121	0.03122
WLS	0.89780	0.01860	0.02921	0.90077	0.02098	0.03058	0.88218	0.02975	0.04344
OLS	0.76253	0.01872	0.02538	0.76240	0.02001	0.07665	0.74381	0.03133	0.03788

Table 7.2 A comparative study of ML, WLS, and OLS for $\hat{\beta}$ in the "errors-in-variables" case when error variances are equal.

	$(\sigma_\epsilon=1;\ \sigma_\delta=1)$			$[\sigma_\epsilon=(0.5x_i+1)/3;\ \sigma_\delta=1]$			$[\sigma_\epsilon=(X_i+8)/9;\ \sigma_\delta=1]$		
	$(\hat{\beta})$	$Var(\hat{\beta})$	$MSE(\hat{\beta})$	$(\hat{\beta})$	$Var(\hat{\beta})$	$MSE(\hat{\beta})$	$(\hat{\beta})$	$Var(\hat{\beta})$	$MSE(\hat{\beta})$
				[k=4; m=10]					
ML	1.00602	0.18830	0.18835	1.05137	0.22038	0.22302	1.08486	0.25747	0.26467
WLS	0.96883	0.02289	0.02386	0.96991	0.02898	0.02988	0.96616	0.02668	0.02784
OLS	0.93881	0.02293	0.02666	0.94007	0.02403	0.02763	0.92181	0.02402	0.03010
				[k=10; m=10]					
ML	0.95708	0.01181	0.01365	1.01259	0.01837	0.01859	1.05883	0.02897	0.03244
WLS	0.92618	0.00893	0.01438	0.90955	0.01442	0.02252	0.92713	0.01069	0.01602
OLS	0.92333	0.00880	0.01468	0.91810	0.00992	0.01664	0.92681	0.01055	0.01588
	$(\sigma_\epsilon=1;\ \sigma_\delta=2)$			$(\sigma_\epsilon=(0.5x_i+1)/3;\ \sigma_\delta=2)$			$(\sigma_\epsilon=(X_i+8)/9;\ \sigma_\delta=2)$		
				[k=4; m=10]					
ML	0.96101	0.13644	0.13803	1.00337	0.13022	0.13022	1.07129	0.17060	0.18019
WLS	0.93271	0.09401	0.09852	0.93110	0.08991	0.09481	0.98108	0.10020	0.10056
OLS	0.94122	0.08478	0.08824	0.93842	0.08553	0.08913	0.96561	0.09386	0.09503
				[k=10; m=10]					
ML	0.98180	0.01909	0.01941	0.99354	0.02515	0.02519	1.00230	0.03117	0.03118
WLS	0.88991	0.01877	0.03087	0.90033	0.02002	0.02996	0.88322	0.02890	0.04259
OLS	0.75986	0.01993	0.07753	0.76366	0.02107	0.07724	0.74296	0.03186	0.09791

CHAPTER 8

CONCLUSIONS

For inferences based on the linear model, estimation in the presence of unequal variances is of prime importance. Throughout the years a number of researchers investigated these estimation problems, employing the Methods of Least Squares, Weighted Least Squares, Maximum Likelihood, or some ad hoc approaches. Another approach is estimation by quadratic functions of the observations, based on sums of squares appearing in the analysis of variance table (e.g., Henderson, 1953; Searle, 1968, 1971). But C. R. Rao (1972) pointed out that "In this method the theoretical basis is not clear, the procedures suggested are ad hoc and much seems to depend on intuition." In a series of four papers, C. R. Rao (J. Amer. Statist. Assoc., 65, (1970), pp. 161-172; J. Multi. Anal., 3, (1971), pp. 257-275; J. Multi. Anal., 4, (1971), pp. 445-456; J. Amer. Statist. Assoc., 67, (1972), pp. 112-115) proposed a new principle called Minimum Norm Quadratic Unbiased Estimation (MINQUE), developed some of its optimality properties and suggested further investigation. This new principle was refreshing and received the interest of researchers in the area; as a result, a number of articles appeared on this topic. However, as pointed out by many authors, there are some drawbacks to this method. The most important weakness of MINQUE is that it may give negative values for estimates of non-negative variances, although J. N. K. Rao (1973) gave some modifications of MINQUE based on intuitive grounds which helped to resolve this difficulty. A second problem is that the MINQUE and Modified MINQU-based estimators do not exist for some models of interest.

The purpose of this dissertation has been to develop the theoretical properties and to compare the performance of the different estimators, e.g., OLS, WLS, ML, MINQUE, Modified MINQUE, for the linear model when the variances are unknown and different. We considered the performance of these above mentioned approaches by means of both asymptotic theoretical results and small samples simulated results. We also developed a new method on the basis of prior likelihoods called Posterior Likelihood (PL) estimation. We compared the properties of PL approach with that of existing approaches. We also presented a comparative study which shows the superiority of the PL technique when suitable prior information is available.

A further approach which has been developed is a preliminary testing framework for choosing the best method and making valid inferences. An empirical study of the use of preliminary tests is also provided. To obtain better estimates, a multiple comparison technique was used for preliminary testing and a further comparative study is provided. We also examined the effects of errors in variables upon the different estimators. An empirical comparison is provided for this case also.

In Chapter 1 a review of the existing techniques for estimating the variance components in the linear model is given together with a detailed discussion of the purpose of this research.

Several theorems on the large sample variances of estimates for regression parameters for both WLS and MINQU-based estimators for normal and non-normal cases are developed in Chapter 2. The main aim behind these theorems is to provide a theoretical basis for comparing the large sample properties of WLS and MINQU-based estimators. Also, the results of these theorems are used for comparison with simulated small sample results in later chapters.

In Chapter 3, a comparison of the asymptotic theoretical results with simulated small sample results is presented. For this comparative study, we considered the model of Jacquez, et al. (Biometrics, 24, (1968), pp. 607-627) and investigated the behavior of WLS and MINQU-based estimators of unequal variances through Monte Carlo study. This study showed that MINQU-based or Modified MINQU-based estimators are not suitable for this problem and that the WLS method provides better estimators. The MINQU- and Modified MINQU-based estimators have the tendency to "stretch out" the estimators and they provide estimates with high variance. Both the theoretical and empirical results show that WLS provides better estimates not only in the normal case but also in non-normal cases for all the different patterns of errors considered. Computer based results for the Cauchy distribution confirm the supermacy of WLS over MINQUE for heavy tailed distribution. OLS estimators are best for very small samples. However, when the number of replicates is small, OLS is to be preferred over the other potentially and more efficient models.

In Chapter 4, we developed a new methodology based on prior likelihoods called Posterior Likelihood (PL) along with its theoretical properties. Some theoretical results about the $\hat{\underline{\beta}}_1$ and its variances are obtained using Gamma prior likelihoods for the regression model. The most striking property of PL is that this method of estimation "shrinks" the variance estimates towards a common value rather than "stretching" them. A comparative empirical study on the basis of OLS, WLS, ML, and PL is presented. The most striking thing about the results for the PL estimators is that, even with different prior likelihoods, the method gives fairly similar and generally quite efficient results. When the parameters of the prior likelihoods are fixed, the asymptotic behavior of the PL estimators is just like WLS.

In Chapter 5, we presented preliminary testing procedures for variance inequality which are then used to select the estimation method used for the regression parameters. A review of the existing techniques for testing equality of variances is also presented. In the search for more efficient estimators we estimate the parameters assuming either a common variance or (all) different variances. To make the choice, we used the Modified Bartlett test and Generalized variance as preliminary tests. On the basis of these adaptive procedures, we obtained estimators which are somewhat more efficient when compared with OLS and WLS estimators. In this overall selective procedure of estimators, there was an open question as to which variances differed. So, in Chapter 6, we tried to identify those variances which may be pooled to further improve the estimators for the regression model, based upon Multiple Comparison methods. We used Fisher's Least Significant test (FSD) as a preliminary test to determine the best estimates of variances. An empirical comparative study is also presented. This study shows that the FSD test as a preliminary test may yield sizeable reductions in the variances and we found that the FSD procedure supercedes the other procedures, e.g., OLS, WLS, generalized variance and Modified Bartlett test. Thus the FSD procedure is recommended as a superior method.

In Chapter 7, we considered the effect of "errors-in-variables" and examined different approaches (theoretically and empirically) to cope with this type of problem. An empirical study of the LS, WLS, and ML methods suggests that (i) maximum likelihood estimator has very little bias but high MSE because of high variability in the errors on the X variables when small samples are used; (ii) when σ_{δ}^2 is small, then the OLS estimates give small bias, but when σ_{δ}^2 is large, then MSE for OLS are still smaller than that for ML unless the number of replications is increased.

Limitation of the Study and Direction for Further Research

1. The asymptotic studies need further development to consider extra terms, particularly to allow for changes in the number of samples as well as the number of replications.
2. As with any numerical studies the range of problems considered is limited, but we feel that the use of the well-established example of Jacquez, et al. (1968) means that the results are likely to be reasonably representative.
3. Preliminary testing procedures for testing the equality of variances need more work to provide for better estimates for the variances of the $\hat{\beta}$. Further work is needed on multiple comparison procedures to establish the best form of test to use and a suitable level of α. Again, better variance estimators must be developed.
4. In the problem of "errors-in-variables" we need further research to establish good estimators for small samples. Possibly further progress could be made by linking the ML and multiple comparison methods.

REFERENCES

Ahrens, H. (1978) MINQUE and ANOVA estimator for one way classification - A risk comparsion. Biometrical J. 20, 535-556.

Ahrens, H., Kleffe, J., and Tensler, R. (1979) Mean squared error comparisons of MINQUE, ANOVA and two alternative estimators under the balanced one way random model. Tech. Rep. P-19/79, Akademic der Wissenschaften der DDR.

Aitken, A. C. (1935) On least squares and linear combination of observations. Proc. Roy. Soc. Edin. 55, 42-47.

Anderson, R. L. and Crump, P. P. (1967) Comparisons of designs and estimation procedures for estimating parameters in a two stage nested process. Technometrics 9, 499-516.

Anderson, R. L. (1975) Designs and estimators for variance components. In: J. N. Srivastava, ed., A Survey of Statistical Design and Linear Models, pp. 1-29.

Anderson, T. W. (1973) Asymptotically efficient estimation of covariance matrices with linear structure. Ann. Statist. 1, 135-141.

Anscombe, F. J. and Tukey, J. W. (1963) The examination and analysis of residuals. Technometrics 5, 141-160.

Bancroft, T. A. (1944) On biases in estimation due to the use of preliminary tests of significance. Ann. Math. Statist. 15, 190-204.

Bancroft, T. A. (1964) Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. Biometrics 20, 427-442.

Bancroft, T. A. and Han, C. P. (1977) Inference based on conditional specification: A note and a bibliography. Internat. Statist. Rev. 45, 117-127.

Barnett, V. D. (1970) Fitting straight lines - The linear functional relationship with replicated observations. Appl. Stat. J. R. Stat. Soc. C 19, 135-144.

Bartlett, M. S. (1937) Properties of sufficiency and statistical tests. Proc. Roy. Statist. Soc. A 160, 268-282.

Bartlett, M. S. and Kendall, D. G. (1946) The statistical analysis of variance heterogeneity and the logarithmic transformation. J. Roy Statist. Soc. B 8, 128-138.

Berhardson, C. S. (1975) Type I error rates when multiple comparison procedures follow a significant F test ANOVA. Biometrics 31, 229-232.

Box, G. E. P. (1953) Non-normality and tests on variances. Biometrika 40, 318-335.

Brown, M. B. and Forsythe, A. B. (1974) Robust tests for the equality of variances. J. Am. Statist. Assoc. 69, 364-367.

Brownlee, K. A. (1953) Industrial Experimentation. Chemical Publishing Co.

Cadwell, J. H. (1953) Approximating to the distributions of measures of dispersion by a power of χ^2. Biometrika 40, 336-346.

Carmer, S. G. and Swanson, M. R. (1973) An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. J. Am. Statist. Assoc. 68, 66-74.

Chaubey, Y. P. and Rao, P. S. R. S. (1976) Modifications and adaptations of the principle of MINQUE. Research Report, University of Rochester.

Cochran, W. G. (1939) The use of the analysis of variance in enumeration by sampling. J. Am. Statist. Assoc. 34, 492-510.

Cochran, W. G. (1941) The distribution of the largest of a set of estimated variances as a fraction of their total. Ann. of Eugen. 11, 47-52.

Cochran, W. G. and Carrol, S. P. (1953) A sampling investigation of the efficiency of weighting inversely as the estimated variance. Biometrics 9, 447-459.

Cochran, W. G. (1968). Errors of measurement in statistics. Technometrics 10, 637-666.

David, H. A. (1952) Upper 5% and 1% points of the maximum F-ratio. Biometrika 39, 422-424.

David, H. A. (1956) The ranking of variances in national populations. J. Am. Statist. Assoc. 51, 621-626.

Dixon, W. J. and Massey, Jr., F. J. (1969) Introduction to Statistical Analysis, 3rd edition. New York: McGraw-Hill Book Co.

Duncan, D. B. (1955) Multiple range and multiple F tests. Biometrics 11, 1-42.

Fairfield Smith, H. (1936) A discriminant function for plant selection. Ann. Eugenics (London) 7, 240-260.

Fisher, R. A. (1920) A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean square error. Monthly Notices of the Royal Astronomical Society, 758-770.

Fisher, R. A. and Yates, F. (1963) Statistical Tables for Biological Agricultural and Medical Research. Hafner Publishing Company, Inc.

Focke, J. and Dewess, G. (1972) Uber die Schatzmethode MINQUE Von C. R. Rao and ihre Verallgemeinerung. Math. Operationforsch. Statist. 3, 129-143.

Fuller, W. A. and Rao, J. N. K. (1978) Estimation for a linear regression model with unknown diagonal covariance matrix. Ann. Statist. 6, 1149-1158.

Gartside, P. S. (1972) A study of methods for comparing several variances. J. Amer. Statist. Assoc. 67, 342-346.

Geng, S., Wang, W. J., and Miller, C. (1979) Small sample size comparisons of tests for homogeneity of variances by Monte Carlo. Comm. Statist. - Simula. Computa. B8(4), 379-389.

Goldman, A. J. and Zelen, M. (1964) Weak generalized inverses and minimum variance linear unbiased estimation. (Unpublished).

Goodnight, J. H. and Hemmerle, W. J. (1978) A simplified algorithm for the W-transformation in variance component estimation. SAS Tech. Rept. R-104, Raleigh, NC.

Halperin, H. (1961) Fitting of straight lines and prediction when both variables are subject to error. J. Am. Statist. Assoc. 56, 657-669.

Harter, H. Leon (1957) Error rates and sample sizes for range tests in multiple comparisons. Biometrics 13, 511-536.

Hartley, H. O. (1950) The maximum F-ratio as a short cut test for heterogeneity of variance. Biometrika 37, 308-312.

Hartley, H. O. (1955) Some recent developments in analysis of variance. Comm. Pure. Appl. Math. 8, 47-72.

Hartley, H. O. and Rao, J. N. K. (1967) Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54, 93-108.

Harville, D. A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Statist. Assoc. 72, 320-340.

Hemmerle, W. J. and Hartley, H. O. (1973) Computing maximum likelihood estimates for the mixed AOV model using the W-transformation. Technometrics 15, 819-831.

Henderson, C. R. (1950) Estimation of genetic parameters. Ann. Math. Statist. 21, 309-310.

Henderson, C. R. (1953) Estimation of variance and covariance components. Biometrics 9, 226-252.

Henderson, C. R. (1977) Prediction of future records. In: Proc. Int. Conf. on Quantitative Genetics, pp. 616-638.

Horn, S. D. and Horn, R. A. (1975) Comparison of estimators of heteroscedastic variances in linear models. J. Am. Statist. Assoc. 70, 872-879.

Horn, S. D., Horn, R. A., and Duncan, D. B. (1975) Estimating heteroscedastic variances in linear models. J. Am. Statist. Assoc. 70, 380-385.

Huntsberger, D. V. (1955) A generalization of a preliminary testing procedure for pooling data. Ann. Math. Stat. 26, 734-743.

Infante, A. (1978) Die MINQUE - Schatzung bei Verlaufskurvenmodellen mit Zufalligen regressionskoeffizienten. Thesis, Dortmud (FRG).

Jacquez, J. A., Mathur, F. J., and Crawford, C. R. (1968) Linear regression with non-constant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. Biometrics 24, 607-627.

Kendall, M. G. and Stuart, A. (1979) The Advanced Theorey of Statistics. London: Griffin, Volume 2, Chapter 29.

Keselman, H. J., Games, P. A., and Clinch, J. J. (1979) Tests for homogeneity of variance. Commun. Statist. - Simula. Computa. B8(2), 113-129.

Keuls, M. (1952) The use of the "Studentized Range" in connection with an analysis of variance, In: Olkin, et al., eds., Contributions to Probability and Statistics. California: Stanford Press, pp. 270-292.

Kitagawa, T. (1963) Estimation after preliminary tests of significance. Univ. of California Publications in Statistics 3, 147-186.

Kleffe, J. (1975) Quadratic Bayes-Schatzungen fur Lineare Parameter: der Kovarianz-matrix im Gemischten Linearen Modellen. Dissertation, Humboldt University, Berlin.

Kleffe, J. (1976) Best quadratic unbiased estimators for variance components in mixed linear models. Sankhya B 38, 179-186.

Kleffe, J. (1977a) Invariant methods for estimating variance components in mixed linear models. Math. Operationforsch. Statist. 8, 233-250.

Kleffe, J. (1977b) A note on ∞-MINQUE in variance covariance components models. Math. Operationforsch. Statist. 8, 337-343.

Kleffe, J. (1978) Simultaneous estimation of expectation and covariance matrix in linear models. Oper. Statist. Ser. Statist. 9, 443-478.

Kleffe, J. (1979) C. R. Rao's MINQUE for replicated and multivariate observations. Tech. Rept. Zimm. der AdW der DDR, Berlin.

Kleffe, J. (1980) C. R. Rao's MINQUE under four two way ANOVA models. Biometrical J. 21, 230-234.

LaMotte, L. R. (1973) Quadratic estimation of variance components. Biometrics 29, 728-730.

Larson, H. J. and Bancroft, T. A. (1963a) Sequential model building for prediction in regression analysis, I. Ann. Math. Stat. 34, 462-479.

Larson, H. J. and Bancroft, T. A. (1963b) Biases in prediction by regression for certain incompletely specified models. Biometrika 50, 391-402.

Layard, M. W. J. (1973) Robust large-sample tests for homogeneity of variances. J. Am. Statist. Assoc. 68, 195-198.

Leslie, R. T. and Brown, B. M. (1965) Use of range in testing heterogeneity of variance. Biometrika 53, 221-227.

Levene, H. (1960) Robust tests for equality of variances. In: Probability and Statistics, I. Olkin, et al., eds. Stanford: Stanford University Press, pp. 278-292.

Levy, K. J. (1975) Some multiple range tests for variances. Educational and Psychological Measurement 35, 817-821.

Lindley, D. V. (1947) Regression lines and the linear functional relationship. J. R. Statist. Soc. Suppl. 9, 219-244.

Mandansky, A. (1959) The fitting of straight lines when both variables are subject to error. J. Am. Statist. Assoc. 54, 173-205.

Miller, J. J. (1977) Asymptotic properties of maximum likelihood estimates in the mixed of analysis of variance. Ann. Statist. 5, 746-762.

Miller, J. J. (1979) Maximum likelihood estimation of variance components - A Monte Carlo study. J. Statist. Comp. and Simulation 8, 175-190.

Miller, G. R. (1966) Simultaneous Statistical Inference. New York: McGraw-Hill.

Miller, Jr., R. G. (1968) Jackknifing variances. Ann. Math. Statist. 39, 567-582.

Miller, R. G. (1977) Developments in multiple comparisons 1966-1976. J. Am. Statist. Assoc. 72, 779-788.

Moran, P. A. P. (1971) Estimating structural and functional relationships. J. Mult. Anal. 1, 232-255.

Newman, D. (1939) The distribution of the range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. Biometrika 31, 20-30.

Neyman, J. and Scott, E. (1948) Consistent estimators based on partially consistent observations. Econometrics 16, 1-32.

Olsen, A., Seely, J., and Birkes, D. (1976) Invariant quadratic unbiased estimation for two variance components. Ann. Statist. 4, 878-890.

O'Neill, R. T. and Wetherill, G. B. (1971) The present state of multiple comparison methods. J. R. Stat. Soc. B 33, 218-241.

Ord, J. K. and Leonard, T. (1976) Investigation of the F-test procedure as an estimation short-cut. J. R. Stat. Soc. B 38, 95-98.

Panse, V. G. (1946) An application of discriminant function for selection in poultry. J. Genetics (London) 47, 242-253.

Patterson, H. D. and Thompson, R. (1975) Maximum likelihood estimation of components of variance. In: Proc. of 8th International Biometric Conference, pp. 197-207.

Powell, M. J. D. (1970) A survey of numerical methods for unconstrained optimization. SIAM Rev. 12, 79-97.

Pringle, R. M. (1974) Some results on the estimation of variance components by MINQUE. J. Am. Statist. Assoc. 69, 987-989.

Pukelsheim, F. (1977) Linear models and convex programs: Unbiased non-negative estimation in variance component models. Tech. Rep. 104, Stanford University.

Pukelsheim, F. (1978a) Examples for unbiased non-negative estimation in variance component models. Tech. Rep. 113, Stanford University.

Rao, C. R. (1947) General methods of analysis for incomplete block designs. J. Am. Statist. Assoc. 42, 541-561.

Rao, C. R. (1953) Discriminant function for genetic differentiation and selection. Sankhya 12, 229-246.

Rao, C. R. (1956) On the recovery of interblock information in varietal trials. Sankhya 17, 105-114.

Rao, C. R. (1970) Estimation of heteroscedastic variances in linear models. J. Am. Statist. Assoc. 65, 161-172.

Rao, C. R. (1971a) Estimation of variance and covariance components. J. Mult. Anal. 1, 257-275.

Rao, C. R. (1971b) Minimum variance quadratic unbiased estimation of variance components. J. Mult. Anal. 1, 445-456.

Rao, C. R. (1972) Estimation of variance and covariance components in linear models. J. Am. Statist. Assoc. 67, 112-115.

Rao, C. R. (1973) Linear Statistical Inference and its Applications, 2nd edition. New York: John Wiley.

Rao, C. R. (1979) Estimation of variance components - MINQUE theory and its relation to ML and MML estimation. Sankhya B 41, 283-297.

Rao, J. N. K. and Subrahmaniam, K. (1971) Combining independent estimators and estimation in linear regression with unequal variances. Biometrics 27, 971-990.

Rao, J. N. K. (1973) On the estimation of heteroscedastic variances. Biometrics 29, 11-24.

Rao, P. S. R. S. (1977) Theory of the MINQUE - A review. Sankhya B, 201-210.

Rao, P. S. R. S. and Chaubey, Y. P. (1978) Three modifications of the principle of the MINQUE. Comm. Statist. Math. A7, 767-778.

Scheffe, H. (1959) The Analysis of Variance. New York: John Wiley.

Searle, S. R. (1968) Another look at Henderson's method of estimating variance components. Biometrics 24, 749-788.

Searle, S. R. (1971) Topics in variance component estimation. Biometrics 27, 1-76.

Seegar, P. (1968) A note on a method for the analysis of significances en masse. Technometrics 10, 586-593.

Seely, J. (1975) An example of inadmissible analysis of variance estimator for a variance component. Biometrika 62, 689-690.

Sinha, B. K. and Wieand, H. S. (1977) MINQUE's of variance and covariance components of certain covariance structures. Indian Statistical Institute Tech. Rep. 28/77.

Smith, W. C. (1974) The combination of statistical tests of significance. Unpublished Ph.D. Thesis, Iowa State University.

Snedecor, G. W. and Cochran, W. G. (1967) Statistical Methods, 6th edition. Ames: Iowa State Press.

Sprent, P. (1966) A generalized least squares approach to linear functional relationships. J. R. Statist. Soc. B 28, 278-297.

Sprent, P. (1969) Models in Regression and Related Topics. London: Methuen.

Stoline, M. R. (1981) The status of multiple comparisons: Simultaneous estimation of all pairwise comparisons in one-way ANOVA designs. The American Statistician 35, 134-141.

Swallow, W. H. and Searle, S. R. (1978) Minimum variance quadratic unbiased estimation of variance components. Technometrics 20, 265-272.

Villegas, C. (1961) Maximum likelihood estimation of a linear functional relationship. Ann. Math. Statist. 32, 1048-1062.

Yates, F. and Zacopancy, I. (1935) The estimation of the efficiency of sampling with special reference to smapling for yield in cereal experiments. J. Agric. Sci. 25, 545-577.

Yates, F. (1940) The recovery of inter-block information in balanced incomplete-block designs. Ann. Eugenics (London) 10, 317-325.

Zellner, A. (1962) An efficient method of estimating seemingly unrelated regression and tests for aggregation bias. J. Am. Statist. Assoc. 57, 348-368.

Zellner, A. and Huang, D. S. (1962) Further properties of efficient estimators of seemingly unrelated regression equations. Int. Econ. Rev. 3, 303-313.